Stem Cell Biology and Clonal Expansion in Normal and Adenomatous Human Intestinal Crypts

Adam Humphries

University College London

and

Cancer Research UK London Research Institute

PhD Supervisor: Professor Sir Nicholas Wright

A thesis submitted for the degree of

Doctor of Philosophy

University College London

September 2011

Declaration

I Adam Humphries confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Abstract

Gastrointestinal cancer is thought to be primarily a disease of stem cells, whereby a tumorigenic stem cell clone can expand within an individual colonic crypt and then within the epithelium to form an adenoma - the pre-malignant lesion of the colon. However, data demonstrating stem cell populations and the dynamics of clonal expansion in human intestinal crypts is lacking. Naturally occurring, somatic clonal mutations in mitochondrial DNA were used to identify the progeny of a putative single stem cell lineage within crypts; this allowed the visualization of putative stem cell clones arising and expanding within human colon and small bowel crypts. Immunohistochemistry for lineage specific markers, to confirm multi-potentiality, and in-situ hybridisation for stem cell markers was then performed to phenotype clones further and to demonstrate a single stem cell lineage and their direct progeny within the human crypt. By combining clonal somatic mutations in mitochondrial DNA with methylation signatures, the dynamics of clonal expansion within normal colon crypts and the epithelium was studied, and using mathematical modeling the time course of these events was able to be estimated. By applying the same techniques, in addition to studying genomic mutations, the dynamics of human adenoma growth was investigated; this suggested that adenomas appear to grow in a punctuated manner, with episodes of rapid clonal expansion being followed by periods of relative quiescence.

Acknowledgements

This work was made possible by the funding and support of Cancer Research UK, in addition to generous donations from the RL St J Harmsworth Memorial Trust and The Jean Shanks Foundation.

I would like to thank all the members of the Histopathology Lab at The London Research Institute, Cancer Research UK, past and present, for their help and guidance throughout my PhD. Specifically, thank you to Rosemary Jeffery and Richard Poulsom for advice and technical assistance with the immunohistochemistry and *in-situ* hybridisation techniques, and Bill Otto for all his support. Thank you to Chung-Yin Lee, for your endless patience and skills with the ordering book and managing the lab budget, and without whom no one would be able to do any work! I am also very grateful for the assistance provide by Emma Nye and George Elia with tissue processing and sectioning, and all the members of the Equipment Park for providing an excellent DNA sequencing service. In addition, I am indebted to Paul Tadrous, for teaching me how to use his image processing software and for being available to help with any trouble-shooting.

I cannot thank Stuart McDonald and Trevor Graham enough for their extensive teaching, technical assistance, intellectual support and close guidance throughout the whole of my PhD. It has been fantastic to work with you both over the last 4 years, and I could not have achieved much of what I have done without your help. I am hugely grateful to Manuel Rodriguez-Justo for his help with patient recruitment and tissue collection, to Marco Novelli for his supervision, advice and pathology expertise, and Simon Leedham for his support throughout my PhD. I would also like to thank Bibek Das, Alice Gutteridge and Daniel Miller for their help with data collection and analysis.

Special mention must go to my beautiful wife, Catriona, and my two sons, Rory and Oscar, without whom there would be little point in doing anything. Finally, to Nick Wright, simply the best supervisor and role model a person could wish for.

Table of Contents

Table of figures

List of tables

Abbreviations

ACF: Aberrant crypt foci

APC: Adenomatous polyposis coli

BGN: Biglycan

BMP: Bone morphogenic protein

BRAF: v-raf murine sarcoma viral oncogene homolog B1

Brdu: Bromodeoxyuridine

CA-II: Carbonic anhydrase II

CBCs: Crypt based columnar cells

CCO: cytochrome *c* oxidase

cDNA: complementary-DNA

CIMP: CpG island methylator phenotype

CIN: Chromosomal instability

CSX: Cardiac-specific homeobox

DAB: Diaminobenzadine

DBA: *Dolichos biflorus* agglutinin

DCAMKL-1: Ca^{2+}/calmodulin-dependent protein kinase

DCC: Deleted in colorectal carcinoma

DEPC: Diethylpyrocarbonate

DSS: Dextran sulphate

EGF: Epidermal growth factor

EGFR: Epidermal growth factor receptor

FAP: Familial Adenomatous Polyposis Coli

FFPE: Formalin fixed paraffin embedded

FZD: Frizzled receptor

G6PD: Glucose-6-phosphate dehydrogenase

H&E: Haematoxylin and eosin

HNPCC: Hereditary Non-Polyposis Colorectal Cancer

HPS: Hyperplastic Polyposis Syndrome

IF: Immunofluorescence

ISH: *In-situ* hybridisation

KI-67: Protein encoded for in humans by the *MK167* gene

KRAS: v-ki-ras2 kirsten rat sarcoma homolog

LCMD: Laser-capture micro-dissection

LGD: Low-grade dysplasia

Lgr5: Leucine-rich-repeat-containing G-protein-coupled receptor-5

LINE-1: Long interspersed nucleotide elements

LOH: Loss of heterozygosity

LRCs: Label-retaining cells

LRP: Lipoprotein receptor-related protein

MALDI-MS: Matrix-assisted laser desorption ionisation mass spectrometry analysis

MAPK: mitogen-activated kinase cascade

MCR: Mutation cluster region

MMR: Mis-match repair

MSI: Microsatellite instability

MSS: Microsatellite stable

MT-CO3: mitochondrial cytochrome *c* oxidase

MT-CYB: mitochondrial cytochrome-B

mtDNA: mitochondrial DNA

mTert: Mouse telomerase reverse transcriptase

MT-ND2: mitochondrial encoded NADH dehydrogenase 2

MT-RNR2: mitochondrial ribosomal RNA 2

MT-TA: mitochondrial transfer RNA alanine

MUC-2: Mucin-2

MYOD1: Myoblast determination protein 1

OAT: O-acetyl transferase

PBS: Phosphate buffered saline

PCR: Polymerase chain reaction

PDGFR: Platelet-derived growth factor receptor

RT-PCR: Reverse-transcriptase polymerase chain reaction

SDH: succinate dehydrogenase

TA: Tubular adenoma

TCF: T-cell factor

TGFBR1: Type-1 transforming growth factor beta receptor

TGF-β: Transforming growth factor beta

TP53: Tumour protein 53

TUNEL: Terminal deoxynucleotidyl dUTP transferase nick end labelling

TVA: Tubulo-villous adenoma

UTP: Uridine tri-phosphate

Wnt: Signalling pathway involving proteins encoded for by family of *WNT* genes

Chapter 1. Introduction

Colon cancer is the third most common cancer in men and the second most common cancer in women worldwide [1], with an individual having an approximate 5-7% lifetime risk of developing the disease in the UK [2], thus contributing significant morbidity and mortality to the general population. This Chapter will discuss the current understanding of how stem cells and their progeny are organized within the intestinal crypt, and the processes that result in tumorigenic mutations becoming fixed within the stem cell niche with subsequent expansion and spread within the epithelium. In the colon the *adenoma* is accepted as the immediate antecedent of colorectal adenocarcinomas, and the multiple histological and genetic steps involved in the evolution of these important pre-malignant lesions will be described.

1.1 The intestinal crypt

The human colonic epithelium consists of a single layer of columnar cells that form the functional unit of the gut – *the crypt* – via finger-like projections into the underlying connective tissue, with a population of approximately 2000 cells per crypt (Figure 1.1). The terminally differentiated epithelial cell types that constitute the epithelium are: (i) Columnar absorptive cells – *colonocytes*; (ii) Mucous-secreting *goblet* cells; (iii) *Endocrine* cells, also known as *neuroendocrine* or *enteroendocrine* cells, that secrete a variety of peptide hormones in an endocrine or paracrine manner and (iv) *Paneth* cells - granular epithelial cells found in abundance at the base of the small intestinal crypt, but only occasionally seen in the colonic epithelium in the ascending colon and some disease states. The true function of Paneth cells is still unclear: They are thought to have an important role in innate immunity and secrete a number of defensins and anti-microbial peptides [3], in addition there is now emerging evidence that they can be very

long-lived and recently it has been suggested that, in the small bowel at least, they may be crucial for supporting the stem cells within the niche at the base of the crypt [4].

The concept that these differentiated cells arise from a single multi-potential stem cell through a number of *committed progenitors* was first formulated by Cheng and Leblond as the *Unitarian Hypothesis* in 1974 [5], and is well founded in animals [6, 7]. The first few generations of stem cell divisions are known as *transit-amplifying* cells, these divide rapidly to significantly increase the number of progeny that result from a single stem cell division, as cells then move up the crypt from the base toward the luminal surface they become committed to a certain lineage - *committed progenitor* cells, as further migration up the crypt axis occurs cells are increasingly differentiated until when they reach the luminal surface terminally differentiation has occurred [8] (Figure 1.1). Once at the luminal surface, differentiated cells are shed 5-7 days after their initial conception in the lower crypt [9].

1.1.1 Multi-potential stem cells are housed within a basal niche

All the experimental evidence to date supports the theory, originally proposed by Williams et al [10], that adult tissue-specific multi-potential stem cells are positioned at the base of the intestinal crypt within a *stem cell niche*. The *niche* is a micro-environment that consists of the stem cells themselves and a sheath of surrounding mesenchymal cells, of the myofibroblast lineage - *pericryptal myofibroblasts*, separated by the basal lamina. Interactions between the pericryptal myofibroblasts and niche stem cells via paracrine secretion of cytokines and growth factors are thought to be important for maintaining the various cell phenotypes. Activation of the Wnt and Notch cascade signalling pathways is required to maintain the undifferentiated 'stem-like' state [11, 12], and, along with factors such as the bone morphogenetic protein (BMP) antagonists gremlin 1 and gremlin 2 – shown to be secreted by the myofibroblasts [13], are involved

in the maintenance of the stem cell habitus and cell migration and differentiation [8, 11, 14] (see section 1.1.4).

However, the extent to which the niche micro-environment, created by the surrounding stromal and mesenchymal cells, plays in maintaining basal crypt stem cells is not clear: Sato et al [15] were able to extract individual crypt stem cells from mouse small intestine and, using only a laminin matrigel and a uniform distribution of a few growth signals, they were able to form self-renewing crypt-villus like structures with stem and Paneth cells located at the base. Thus, the intrinsic (epi)genetic programming within the stem cells themselves may well determine the differential response of crypt cells to external signalling pathways, such as Wnt and Notch. More recently, studies in mice have also suggested a potential role for basal Paneth cells in supporting intestinal stem cells within the niche via the expression of Wnt and Notch ligands [4].

1.1.2 Identifying intestinal stem cells and their niche – *self-renewal and multipotency*

Much evidence for the structure of the intestinal crypt has come from murine labeling experiments using somatic mutagens or chimeric mice, where one parent strain carries a demonstrable marker. Most of the work has been done on the small intestine, although the principles have been applied to the colon as many similarities exist. Debate continues to exist over the exact location of the niche within the crypt, numbers of stem cells and the properties of niche dynamics.

Evidence that the intestinal crypt is supported by a number of stem cells residing within a niche, rather than a single stem cell as initially thought [16], first came from a mouse study by Williams et al [10]. They marked cells via mutagen-induced loss of the X-linked enzyme glucose-6-phosphate dehydrogenase (G6PD) in crypt cells, and compared the time-course of mutated phenotypes in both the small and large intestine. A short time

after administration of the mutagen a high frequency of partially-mutated crypts were observed, over time a rapid increase in wholly-mutated crypts was observed that reached a plateau at the same time that partially-mutated crypts disappeared. The time taken for the appearance of wholly mutated crypts in the large intestine was approximately 4 weeks, but 21 weeks in the small intestine. The authors concluded that their results were consistent with several stem cells supporting a crypt niche, and that the number of stem cells may differ between the small and large colon, with a greater number of niche stem cells in the small intestine to explain the time differences observed. The hypothesis that the intestinal crypt is maintained by a stem cell niche housing multiple related stem cells has since been supported by further studies in both the mouse and human [17-20].

It is not possible to morphologically differentiate stem cells from other cells within the crypt, however there have been a number of markers which have been previously proposed as markers of stem cells in the intestine: these include molecules such as *Musashi-1*, an RNA-binding protein thought to be involved in stem cell divisions during *Drosophila* neural development [21, 22], and DCAMKL-1, a Ca (2+)/calmodulin (CaM)-dependent protein kinase [23]. The evidence for such claims is limited, being based mainly on the finding of a small number of cells labelled with the marker in the putative stem cell position towards the base of the crypt. However, location is not a marker of stemness and recent experimental approaches have used the unique ability of stem cells, once clonally-marked, to re-populate the crypt with clonal descendants in all cell lineages - *clonal replacement* or *monoclonal conversion* (Figure 1.2C), to demonstrate the two key properties that define stem cells: Self-renewal and multi-potency. Thus Barker et al [19] were able to demonstrate that *Lgr5* (leucine-rich-repeat-containing G-protein-coupled receptor 5, also known as *Gpr49*), an intestinal Wnt target gene, showed exclusive expression in cycling crypt base columnar cells (CBCs), previously identified many years ago as potential stem cells at the very base of the murine small intestinal crypt between the Paneth cells [5, 24, 25]. Additionally, and very importantly, they used an inducible *Cre* knock-in allele under the control of the *Lgr5* promoter, and the *Rosa26-lacZ* reporter strain to irreversibly label *Lgr5*-expressing cells and demonstrated

that the Lgr5-positive ($Lgr5^{+}$) CBCs generated all epithelial lineages over a 60-day period. Thus *clonal replacement* or *monoclonal conversion* is a very powerful tool for the identification of stem cells, the location of stem cell niches, and for the analysis of lineage relationships in the gut and elsewhere.

In the Barker et al [19] Lgr5 study expression was limited to 6-8 CBCs located in-between the basal Paneth cells in the small intestinal crypt, and a few CBCs at the bottom of the colonic crypt. Previous DNA-label retention studies [26, 27], proposed as surrogate markers of stem cells, had identified the stem cell niche in the small intestine as being above the basal Paneth cells, around cell position +4 – the "+4 position model" (Figure 1.1A). Thus the findings of Barker et al [19] argued against this, and was in favour of the "stem cell zone" model proposed by Leblond and colleagues [5, 24] (Figure 1.1A). Hans Clevers and colleagues then performed further experiments studying the clonogenic capacity of Lgr5+ CBCs [15]: single Lgr5+ cells isolated from mouse small intestine were able to generate long-lived organised crypt-villus structures in culture that contained all differentiated cell types, further supporting *Lgr5* expression as a specific marker of intestinal stem cells. They were able to grow and maintain the crypt-villus structures using a culture system containing only a laminin based matrigel, epidermal growth factor (EGF), the Wnt agonist R-Spondin 1 and the BMP inhibitor noggin, and concluded that a non-epithelial cellular niche is not required for the formation of an organised intestinal crypt. The same group then further challenged the role of the peri-cryptal myofibroblasts in maintaining stem cells, by providing evidence that small intestinal Paneth cells and basal Paneth-like cells in the mouse colon produce Wnt and other niche signals and are essential for the maintenance of intestinal stem cells [4].

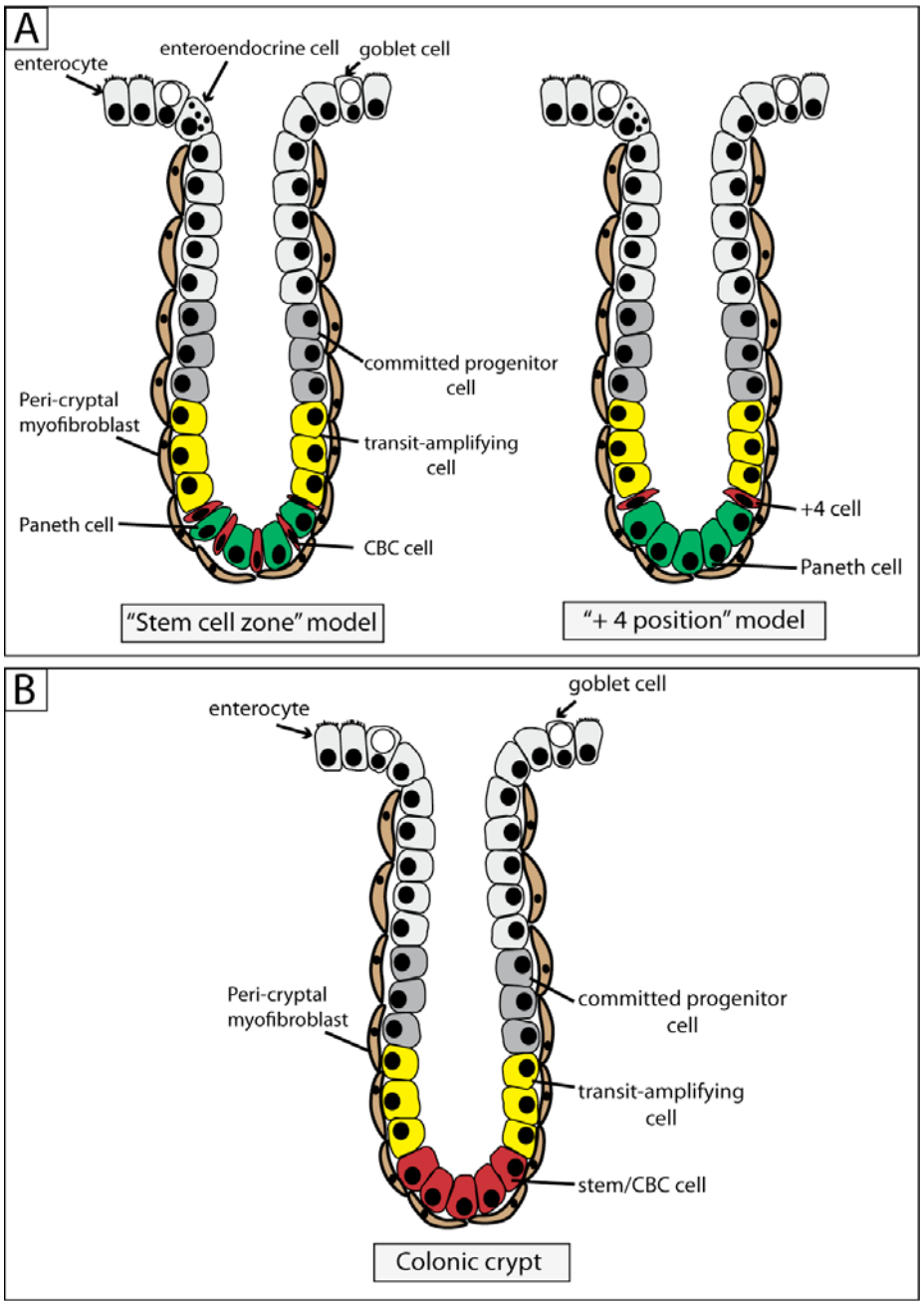

Figure 1.1 Small and large bowel crypt structure
(A) Small intestine: In the "Stem cell zone model" the stem cells are the actively cycling crypt based columnar cells (CBCs) located between the basal Paneth cells, surrounding myofibroblasts comprise the 'niche'; the stem cells are located above the Paneth cells, at cell position 4, in the "+4 position" model. Stem cell divisions produce the transit amplifying cells, which rapidly proliferate; as cells migrate up the crypt they become committed to either a secretory or absorptive lineage and are terminally differentiated once they reach the top. **(B)** Large intestine: Less work has focused specifically on the large bowel, but the evidence points to a small number of stem cells located at the base of the crypt [19]; recently 'Paneth-like' CD24+ cells have been identified in-between the basal CBCs in the mouse colon [4].

The cycling of stem cells has been an area of much controversy: It has been proposed that, in order to protect their DNA, stem cells rarely cycle [26], and this was a property that many long-term DNA label retention studies used in order to identify putative stem cells in the +4 position of the small intestinal crypt [26, 27] – termed *label retaining cells* (LRCs). However the CBCs identified by Barker et al [19] in the mouse were actively cycling, with bromodeoxyuridine (Brdu) labelling of Lgr5+ cells suggesting an average cycling time of one day in the small bowel. A second study by Sangiorgi and Capecchi [20] used a similar mouse genetic construct to permanently label cells expressing the gene *Bmi1*, known to be involved in self-renewal of haematopoietic and neuronal stem cells, and their progeny. *Bmi1*-expressing cells were predominantly seen around the +4 position in the proximal small intestine and were able to self-renew and generate all cell lineages. Interestingly, the Bmi1+ cells did not appear to be as actively cycling as those in the Barker et al [19] Lgr5 study, and were only found in the duodenum and jejunum. However, it may be that these two putative stem cell populations overlap, as Lgr5+ cells also express the highest levels of Bmi1 [28]. A recent study by Montgomery et al [29] demonstrated that expression of mouse telomerase reverse transcriptase (*mTert*) marks slowly cycling cells, in a distribution similar to LRCs - predominantly around cell position +4, that give rise to all differentiated cell types and are distinct to the $Lgr5^{+}$ population. Interestingly, the *mTert*-expressing putative stem cells co-expressed Bmi1 and have previously been shown to also mark long-term Brdu-retaining cells in the small intestinal crypt [30]. Thus, there is now evidence for active and quiescent populations of stem cells co-existing in the intestinal niche, as has been shown for other mammalian tissues [31].

A series of experiments by Potten and colleagues [27, 32, 33] utilised the sensitivity of stem cells to very small doses of radiation to reduce crypt stem cell numbers to differing extents; subsequent measurements of apoptotic cells and the regenerative ability of the crypt enabled estimates of the numbers and location of clonogenic cells in the crypt. This led the authors to conclude that, in the mouse small intestine, each crypt consists of approximately 300-450 cells with 4-6 ultimate–lineage ancestor stem cells in the +4 position, above the Paneth cells. A hierarchical stem cell structure within the crypt was

proposed, with a further two tiers of transit-amplifying cells above the ancestor stem cells retaining clonogenic stem cell properties and having the ability to repopulate the crypt epithelium and regenerate the lower tier stem cells if the lower cells are damaged. In the large intestine the results were broadly similar but a differing regional distribution of the putative stem cell niche was noted. The caecum appeared more similar to the small bowel, however the highest levels of apoptosis were seen at the base of the crypt in the mid-colon and rectum. Numbers of clonogenic cells estimated for the large intestine varied from 5-36, depending on the radiation dose used.

Thus the differing locations and dynamics of putative stem cells detected by authors in the murine small bowel may be due to the fact that overlapping heterogeneous sub-populations of stem cells are being identified by the respective stem cell markers employed. There also appears to be significant differences between both the small intestine and proximal and distal colon.

1.1.3 Stem cell dynamics – niche succession

Results accrued from several tissues in a number of organisms indicate the existence of stem cell niches [34-37], each containing a small number of self-renewing stem cells. There is good evidence that in some niches, such as the Drosophila ovary and testis, stem cells undergo asymmetric cell divisions, giving rise to one stem cell and one cell committed to differentiation [38, 39], this is termed a *lineage mechanism* or *intrinsic asymmetry* [34, 40]. It was previously argued that the majority of stem cell divisions in the intestine were asymmetric, with the stem cell always retaining the template strand of DNA as this would protect against the accumulation of replication errors – termed the *Cairns* or *immortal strand* hypothesis [41].

However, stem cells can also divide symmetrically, producing two cells committed to differentiation or two stem cells, enabling gain or loss of one stem cell lineage from the niche [34] (Figure 1.2A). More recently, evidence from the mouse points to stem cell homeostasis within the intestinal niche exhibiting what is termed *population asymmetry* [37, 42, 43], whereby stem cells are either lost or gained via symmetrical divisions on an equal basis, in order to maintain homeostasis and a constant stem cell population. Snippert et al [43] followed the fates of clonally labelled Lgr5+ cells by generating a multi-colour cre-reporter mouse that allowed for the identification and distinction of individual Lgr5+ stem cells within the niche. Short and long-term clonal tracing data was consistent with the majority of stem cell divisions being symmetrical, with neutral competition between stem cells due to space and number restriction within the niche resulting in neutral drift dynamics, whereby clones expand and contract randomly until one stem cell lineage inevitably takes over the niche [42] (Figure 1.2B). Of course, there may be other stem cell populations within the crypts, such as those marked by *Bmi1* and *mTert* expression, that have different dynamics with a higher proportion of asymmetric divisions.

1.1.4 Methylation patterns enable insights into the human intestinal niche

It is by the process of symmetrical divisions, in addition to random apoptosis of stem cells, that one stem cell lineage can, stochastically by neutral drift or by a selective advantage incurred as a result of a mutation, come to dominate a niche, a process known as *niche* or *clonal succession* (Figure 1.2B). The progeny of this stem cell will then go on to occupy the whole crypt – termed *monoclonal conversion* (Figure 2C). The approach by Shibata and colleagues has given considerable insights into the existence of a stem cell niche and the mechanisms of niche succession in the human colon [18, 44, 45]. Realising that the detection of a stem cell niche has a requirement to distinguish between cells, and given the morphological identity between basal crypt cells in the human, they proposed that stem cell lineages might be distinguished from each other by the methylation status of CpG islands within the coding sequence of non-expressed genes, known to increase with age in normal human colon [46, 47]. Methylation at CpG loci

exhibits somatic inheritance and they become polymorphic with ageing [48, 49], therefore cell populations or stem cell lineages that are closely related should share similarities in these epigenetic sequences, as they are not under any selective pressures. The rate of change of methylation status (methylation or de-methylation) at these CpG sites is low, and has been estimated at approximately 2×10^{-5} per CpG site per division [18, 44], thus it is possible to re-create cell histories by comparing these epigenetic tags between cell populations as they act as a molecular clock. Studying methylation patterns of such CpG islands in non-expressed genes in the colon, they concluded that, due to the variance of methylation tags within crypts, age-related methylation is consistent with human crypts being maintained by niches containing *multiple related stem cells*, rather than by immortal stem cells. The data indicated that *successive niche succession cycles* take place in human colonic crypts: a natural consequence of the existence of symmetric stem cell divisions and neutral drift within the niche [37], and the authors estimated that after approximately every 8 years all stem cell lineages within a crypt niche except for one become extinct - some 9 to 10 times in a lifetime [44]. Thus *clonal succession is an inherent characteristic of a niche.* The appreciation that clonal succession can occur in a stem cell niche in the absence of mutation is a crucial one: even if early genetic changes, critical for tumorigenesis, do not give growth advantages, such mutations might 'hitchhike' to dominate the niche through such niche succession cycles [50].

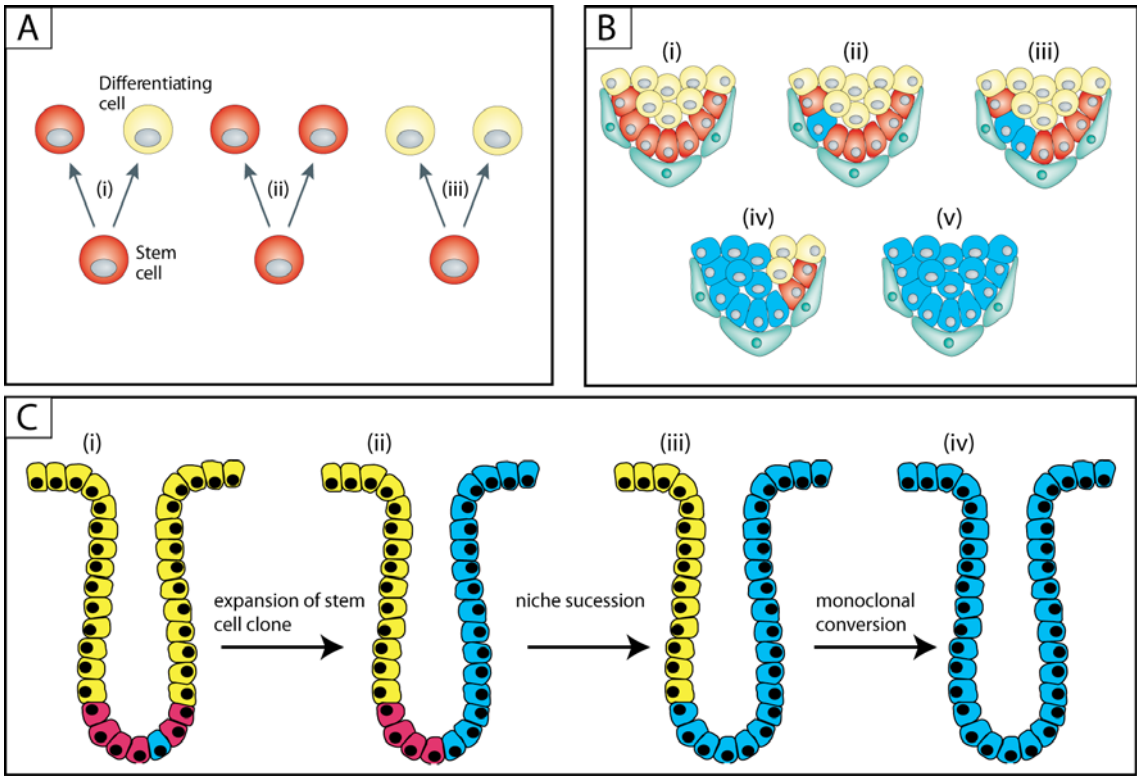

Figure 1.2 Clonal succession is inherent to the niche
(A) Stem cell divisions: (i) an asymmetric stem cell division yields two daughter cells, one of which remains within the niche as a stem cell and the other leaving the niche being destined for differentiation, these are also know as *q* divisions [51]; symmetrical stem cell divisions either produce two stem cells that remain within the niche – *p* divisions (ii), or two daughter cells that exit the niche and differentiate – *r* divisions (iii), with the resulting extinction of that stem cell line [44, 52]. **(B)** Niche succession: most stem cell divisions within the intestinal niche are now thought to be symmetrical [37], as the size and number of stem cells within the niche is tightly regulated, neutral drift results in the process of *niche succession*, whereby one stem cell lineage will stochastic expand to take over the niche. **(C)** Monoclonal conversion: the progeny of a stem cell that has expanded with the niche will then take over the whole crypt.

1.1.5 Wnt signalling, lineage progenitors & determinants of cell fate

There are a number of key regulatory signal pathways that are thought to be important in organizing the architecture of the crypt, maintaining the stem cell phenotype and controlling cell migration and differentiation. Although most of the work done on evaluating these pathways has been done on the small intestine, similar control mechanisms are thought to operate in the colon.

Wnt glycoproteins are a large family of secreted growth factors that produce a wide range of biological effects by binding to their corresponding receptor families frizzled (FZD1-8) and low-density lipoprotein receptor-related protein (LRP5 or LRP6) [53-55]. Canonical Wnt signalling (β-catenin/T-cell factor [TCF] dependent) is restricted to epithelial cells [56] and high levels of Wnt signaling are found at the crypt base, the location of the stem cell niche, and this is thought to be crucial for maintaining cells in the proliferative and undifferentiated 'stem-like' state [8, 57]. In the absence of canonical Wnt ligand a complex consisting of the Adenomatous Polyposis Coli (APC) protein, axin, casein kinase I and glycogen synthase kinase 3β induces phosphorylation of the N-terminal of β-catenin, targeting it for ubiquitin-proteosome mediated destruction [58, 59]. Once Wnt ligands bind and activate the FZD/LRP receptors, the β-catenin destruction complex is inactivated and β-catenin translocates to the nucleus where it binds TCF/Lymphoid enhancer factor (LEF) transcription factors and drives the expression of Wnt target genes [8, 57].

Notch genes encode trans-membrane receptors that are important in regulating communication between cells, and, in combination with Wnt, Notch signalling is also crucial for maintaining proliferation and determining lineage commitment within the crypt [57]. Inhibition of Notch results in a large loss of proliferative crypt cells, and both Wnt and Notch pathway activation appears to be required in combination in order to maintain the stem cell phenotype at the crypt base [11]. However, activation of the bone morphogenic protein (BMP) pathway, involving long and short-range signalling

between mesenchymal and epithelial cells, serves to antagonise crypt formation, suppresses Wnt signalling and inhibits stem cell renewal [60-62].

Wnt signalling is thought to have a variety of important roles, being not just essential for cellular proliferation but also having a role in determining cell fate and the physical location of cells within the crypt. Ablation of the *β-catenin* gene in mice results in the rapid (days) loss of proliferating cells and crypt structures [12], and *Tcf4* null mice have intestinal crypts consisting of only differentiated goblet and absorptive cells [63]. Mutations in the adenomatous polyposis coli *(Apc)* gene result in over-activation of Wnt signalling and formation of multiple polyps that consist of huge numbers of giant crypts that grow without limit [64]. Wnt signalling has also been shown to control the expression of the EphB2 and EphB3 receptors (EphB) and their ligands, Ephrin-B1 and Ephrin-B2, that are involved in maintaining cellular boundaries [14, 65] . Wnt switches on expression of EphB2 and EphB3 receptors and inhibits Ephrin B1/B2 production, this restricts these cells to the crypt base; as one moves up the crypt axis there is a gradient of increased Ephrin ligand expression and loss of Wnt signalling/EphB receptor expression that serves to prevent downward migration of cells due to repulsion between EphB receptors and Ephrin-ligands. EphB2 and EphB3 null mice have proliferative and differentiated cells intermingled throughout the crypt, and selective deletion of EphB3 results in Paneth cells scattered throughout the crypt-villus axis [14].

Blocking of Wnt signalling results in the loss of secretory cell lineages, but enterocyte differentiation appears to be unaffected [65], and excessive Wnt activation in mice promotes Paneth cell differentiation via transcriptional effects of β-catenin/Tcf-4 on specific Paneth cell markers [66]. Notch signalling also has an inhibitory effect on secretory differentiation and it is thought that cells escaping Notch inhibition are able to commit to a secretory lineage, production of Notch Delta/Jagged ligands by these cells then results in lateral inhibition of neighbouring cells from following the same lineage pathway [8]. Math1 was first identified as a transcription factor important for the differentiation of neuronal cells [67], and has since been shown to be expressed in the gut

epithelium with Math1 protein being required for secretory cell differentiation [68]. Hes1 is a Notch signalling component and a negative regulator of secretory cell differentiation - deletion of Hes1 results in increased numbers of secretory cell types and elevated Math1 expression [69].

Thus, in summary, Wnt signalling drives the expression of Notch pathway components, those cells that become committed to the secretory lineage escape Notch activation via expressing the Delta/Jagged ligands with subsequent inhibition of Hes1 and increased Math1 expression ($Wnt^+/Notch^-$); $Wnt^+/Notch^+$ cells continue to proliferate to produce undifferentiated daughter cells, however the pool of Wnt^+ cells is limited, thus some are forced out and lose Wnt ($Wnt^-/Notch^+$), under lateral inhibition from committed secretory progenitors, activated Notch signalling up-regulates Hes1 with resulting suppression of Math1 and commitment to an absorptive fate.

Little is currently known about the intermediate progenitor cell types in the crypt, Bjerknes et al [17] used chemical mutagenesis to randomly mark crypt cells by somatic mutation of the *Dib-1* locus in SWR and F1 mice, thus they were able to identify and study the dynamics of mutant clones containing only mucus cells, only columnar cells and clones that contained a mixture of cell types. The authors concluded that both short and long-lived mucus and columnar cell progenitors exist in the crypt, with long-lived progenitors located in the lower third of the crypt and short-lived progenitors seen higher up the crypt axis. In this study, over 90% of persisting mutant clones containing a mixture of cell types were seen to arise from a crypt base columnar cell in the colon – supporting this as the location for the stem cell.

1.1.6 Intestinal crypts are clonal populations

There is a large body of evidence to support the theory that adult individual crypts are monoclonal, containing cell populations ultimately derived from a single multipotential stem cell, and we have already discussed how *niche succession* with subsequent *monoclonal conversion* is inherent to the dynamics of a stem cell niche.

Studies utilizing mouse chimeras have given valuable insights into the clonal evolution of intestinal crypts [10, 17, 70, 71]. Utilising differential binding of the lectin *Dolichos biflorus* agglutinin (DBA) to B6-derived and not SWR-derived cells, Schmidt et al [70] created C57BL/6J Lac (B6) ↔ SWR mouse aggregation chimeras in order to assess crypt clonality in the developing mouse small intestine. Mixed crypts containing cells of both genotypes were seen in the neonatal period, but by day 14 these had all disappeared and only crypts containing cells of a single genotype were observed. Thus, intestinal crypts are polyclonal at birth but by day 14 all crypts are monoclonal, with crypts undergoing an apparent purification during the rapid growth and high rates of crypt fission occurring at this time. Further mutagenic studies to randomly mark crypt cells observed the disappearance of partially-mutated crypts occurring at the same time as the appearance of the wholly-mutated phenotype, consistent with clonal purification [17, 71, 72].

XO/XY mosaicism is very rare in humans, yet these cases do permit clonal analysis of human crypts. In a study by Novelli et al [73] non-isotopic *in-situ* hybridization studies for the Y chromosome in an individual who had undergone a prophylactic colectomy for Familial Adenomatous Polyposis Coli (FAP), and was also an XO/XY mosaic, demonstrated that colonic crypts were comprised exclusively of either XO or XY cells. Thus, crypts are clonal and derived from a multipotential stem cell. Additional studies on human tissue in individuals heterozygous for naturally occurring polymorphisms of the glucose-6-phosphate dehydrogenase (G6PD) Mediterranean mutation (563 C>T)

and the gene encoding for the enzyme O-acetyl transferase (OAT), have confirmed the conclusion that human intestinal crypts are clonal populations [74, 75].

Recently it has been shown that intestinal stem cells accrue non-pathogenic mutations in their mitochondrial DNA (mtDNA), specifically mutations causing a deficiency in cytochrome *c* oxidase (CCO) - a component of complex IV of the respiratory chain, that are relatively common and increase with age [76-79]. These mutations are then passed on to their progeny. The mitochondrial genome is prone to mutation due to high oxidative stress, a lack of protective histones and poor DNA repair mechanisms. Non-pathogenic mtDNA mutations can expand stochastically within a stem cell by genetic drift, and over time cells will become either *homoplasmic* – all the mitochondria in the cell are mutated, or *heteroplasmic* – the cell contains a mixture of mutated and wild-type mitochondria [80]. This stochastic expansion is a lengthy process, and for a mutated cellular phenotype to be observed homoplasmy or a high degree of heteroplasmy must be present [81], thus stem cells are the only cells that have a sufficient life-span to accumulate these mitochondrial mutations to a level that results in a biochemical deficiency. Two-colour enzyme histochemistry can be used to simultaneously detect the mtDNA-encoded cytochrome *c* oxidase and nuclear DNA-encoded succinate dehydrogenase (SDH), a component of complex II of the respiratory chain. Three different types of crypts are observed: wild-type brown crypts, wholly CCO-deficient crypts and partially CCO-deficient crypts containing both CCO-deficient and wild-type cells. Laser micro-dissection on wholly CCO-deficient crypts followed by PCR coupled with mtDNA sequencing confirm the presence of conserved clonal mitochondrial mutations within colonic crypts [76, 77], confirming that crypts are clonal populations, as here a crypt has been repopulated from a single mutated stem cell via *niche succession* and *monoclonal conversion*. There is no evidence for a selective advantage offered by these mtDNA mutations: Studies using CCO-deficiency as a marker of clonal populations in human stomach and small intestinal have shown CCO-deficient glands or crypts to appear no different to neighbouring crypts [78, 79], when examining for differentiated cell types and proliferative indices. A study by Noteboom et al [82] did

show evidence of a small, but significant decrease in proliferation as assessed by Ki-67, and increase in apoptosis in CCO-deficient human colon crypts versus wild-type.

The time taken for a stem cell lineage to be either lost or become dominant within the niche is known as the *clonal stabilization time* [74]. We have already discussed that this has been shown to differ between the small and large intestine in the mouse, being only 4 weeks in the colon as opposed to 12-20 weeks in the small bowel [10, 71]. Campbell et al [74], by using loss of O-acetylation of sialomucins after pelvic irradiation in patients heterozygous for O-acetyltransferase gene activity as a clonal marker, estimated the clonal stabilization time in humans to be approximately one year; this was supported by modelling studies using methylation patterns of CpG islands that gave an estimated clonal stabilisation time of 220 days in human crypts [18].

In summary, within the human intestinal crypt, multiple stem cells reside within a basal niche – possibly greater than 8 in number; stem cells predominantly undergo symmetric divisions to produce daughter cells that leave the crypt and form the population of transit-amplifying cells. Cells then become committed to a certain lineage, and in the lower part of the crypt a population of long-lived committed precursors exist, with their progeny subsequently undergoing further differentiation and migrating to the luminal surface. The exact life-span of committed progenitor cells in the human crypt is unknown, although it is thought this could be many years. Within the niche, the processes of niche succession and monoclonal conversion enable a single stem cell line and its progeny to completely re-populate the crypt, with successive niche succession cycles taking place over the life-time of a crypt.

1.2 The histological progression to an adenoma

The adenoma-carcinoma sequence is well established as the pathway for colorectal tumorigenesis (see section 1.4), in this section the initiation of the earliest detectable dysplastic lesion – the *monocryptal adenoma*, and their subsequent progression to *microadenomas* and a macroscopically visible adenoma will be discussed (Figure 1.4).

1.2.1 The monocryptal adenoma

The basic cellular hierarchical structure of the colonic crypt consists of a basal stem cell niche that has niche succession and monoclonal conversion as inherent properties, and it is via this process that mutations in stem cells that confer no selective advantage can come to dominant the niche and, subsequently, their progeny occupy the whole crypt. Now consider a somatic mutation in the tumour suppressor gene *APC*, which is thought to be the initial genetic change in most colorectal adenomas, resulting in a niche stem cell that is *APC*+/-. Random genetic drift or natural selection - if haplo-insufficiency confers an advantage - may lead to the mutated stem cell becoming fixed within the crypt [44]. *APC* encodes a large, multifunctional cytoplasmic protein that binds and down-regulates β-catenin – vital in the regulation of Wnt signalling, maintenance of apoptosis, cell cycle progression and chromosomal stability [83], and mutations in this gene can be identified in up to 80% sporadic colorectal carcinomas [84]. Although crypts filled with *APC*+/- cells appear normal [44], there is evidence that haplo-insufficiency confers an increased ability of these crypts to divide and an expansion of the crypt basal cell population [83, 85]. Thus, *APC*+/- crypts are able to divide and expand faster than wild-type crypts, with a resulting field of mutated crypts with increased numbers of stem cells at risk for a further, crucial mutation in *APC*.

The loss of both alleles for *APC* is thought to be sufficient for the development of an adenoma [86]. Once the second *APC* allele is lost, niche succession and monoclonal

conversion results in a dysplastic *APC*-/- crypt surrounded by otherwise histologically normal tissue – the *monocryptal adenoma* [87]. Monocryptal adenomas are the earliest detectable precursors in the adenoma-carcinoma sequence, they frequently occur in mucosa of FAP patients [73] and can also be seen in sporadic cases [88]. These lesions are clonal [73], as would be expected from our knowledge of the monoclonality of individual colonic crypts.

1.2.2 Laying the foundations – *crypt fission and field cancerisation*

Slaughter et al. [89] originally proposed the term *field cancerisation* to explain the presence of multifocal head and neck cancers developing out of a field of precancerous change that had developed as a consequence of carcinogen exposure. The theory was further expanded by Braakhuis et al [90], who proposed that the field was in fact a clonally-expanded area of mutated cells. Thus, the field cancerisation hypothesis [91] requires the generation of a field lesion of genetically altered cells as a specific stage in epithelial carcinogenesis. There has been debate over the next stage to account for field cancerisation in the intestine, and the progression of a monocryptal adenoma to an established adenoma: whether dysplastic cells spill over to invade and colonize the territory of adjacent crypts – a *top-down* process [92], or whether the monocryptal adenoma spreads by *crypt fission*, whereby the crypt divides from its base to form two daughter crypts – the *bottom-up* proposal [93] (Figure 1.3). While it is difficult to obtain definitive dynamic proof in humans, clonal expansion by crypt fission appears to be the prominent mode of spread of an adenomatous crypt [94], and is now generally accepted as the most important mechanism for clonal expansion in early adenomas [95] (Figure 1.3). In conditions where there is mucosal damage or injury, such as ulcerative colitis, repair by epithelial restitution with migration, proliferation and differentiation of epithelial cells to form new crypts, may also offer a chance for mutated clones to develop and expand. However, in ulcerative colitis increased rates of crypt fission are observed [96], therefore crypt re-population and epithelial repair may also occur by crypt fission of neighbouring tissue, and so the contribution of epithelial restitution to clonal expansion and field cancerisation is unclear.

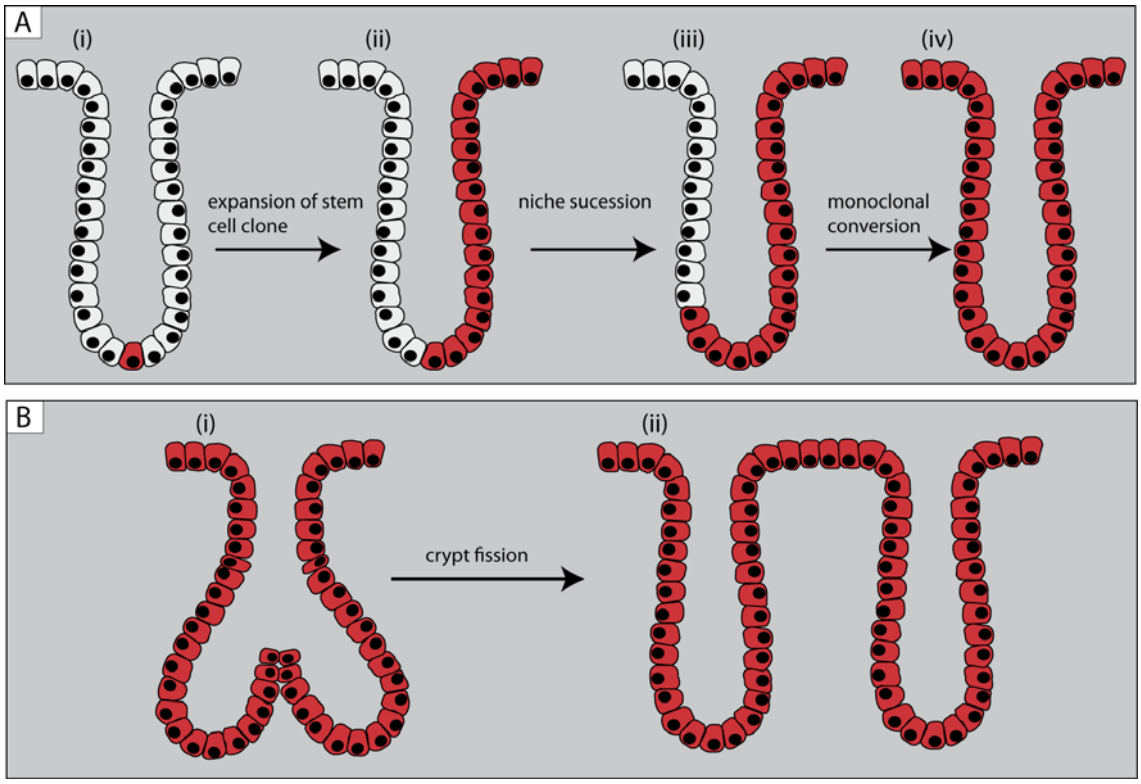

Figure 1.3 Expansion of a mutated stem cell within the colonic crypt and epithelium
(A): a mutated stem cell (highlighted in red) within the niche is able to expand via niche succession (i-iii), subsequently all the progeny of that stem cell lineage take over the crypt – monoclonal conversion, resulting in a monocryptal adenoma (iv). **(B):** The mutated crypt then clonally expands within the epithelium by crypt fission (i), forming an aberrant crypt focus or microadenoma (ii).

1.2.3 Building the lesion: The microadenoma

The next histological step that is observed in the formation of an adenomatous polyp is the development of a *microadenoma* or *aberrant crypt foci* (ACF) – microscopic clusters of small numbers of abnormal crypts (Figure 1.4); these can be divided into two types: (i) non-dysplastic and (ii) dysplastic.

Non-dysplastic ACF are associated with the formation of hyperplastic polyps [97, 98] and appear to be initiated by mutations in the v-Ki-ras2 Kirsten rat sarcoma viral oncogene homolog (*KRAS*) gene; the pre-malignant potential and significance of non-dysplastic ACF and hyperplastic polyps in colorectal carcinogenesis is unclear. The serrated adenoma pathway has been proposed as an alternative model for the development of a small proportion of sporadic colorectal cancers [99], here epigenetic silencing of non-dysplastic ACF, via methylation of genes involved in senescence, apoptosis and DNA repair, is thought to enable the initiation of hyperplastic polyps that then may progress to serrated adenomas with malignant potential (see section 1.4.2).

Dysplastic ACF can be identified in both FAP and non-FAP patients and are likely to represent the histological progression of a monocryptal adenoma to a microadenoma. Genetic studies of these lesions in both FAP and non-FAP patients [98, 100] have shown that in FAP patients all dysplastic ACF contain somatic mutations in *APC*; in dysplastic sporadic ACF *APC* mutations are infrequently found, but mutations in *KRAS* are common (68% in the study by Takayama et al [98]). These findings suggest an alternative genetic pathway in the initiation of sporadic adenomas that does not involve *APC* mutations as the primary event, although somatic mutations in *APC* were detected in a high proportion of established adenomas in the same patients.

This is figure is not available due to third party copyright restrictions

Figure 1.4 The histological progression to a visible adenoma
Normal colonic mucosa shown in **(A)**. In **(B)** a dysplastic crypt is seen surrounded by normal crypts – a monocryptal adenoma (image reproduced from Preston et al [93]). Expansion of a dysplastic crypt, thought to be by crypt fission, forms an aberrant crypt focus or microadenoma **(C)** (image reproduced, from Redston [101]), further growth then results in an established adenoma **(D)**.

1.2.4 Established adenomas are polyclonal populations

If one assumes that a single stem cell undergoes neoplastic transformation and subsequent niche succession with clonal conversion forms a monocryptal adenoma that expands by crypt fission, we would expect early adenomas to be monoclonal. Studies in human females using patterns of inactivation of X chromosome-linked genes as a clonal marker have, indeed, concluded that adenomas are monoclonal in origin [102-104]. However, these indirect observations are heavily biased towards showing monoclonality due to the large X-linked patch size in humans [75].

Direct clonal analysis of adenomas in an XO/XY individual with FAP demonstrated that 76% of adenomas were polyclonal [73], and this data was supported by a study in $APC^{min/+}$ mice chimeric for the *ROSA26* cell lineage marker [105] that showed normal intestinal crypts to be monoclonal, but estimated that 79% of adenomas had an overtly polyclonal structure, even at an early stage. Shibata et al [106], were able to identify at least two distinct *KRAS*-mutated sub-clones in four out of seven adenomas studied, indicating that a high proportion of adenomas demonstrate heterogeneity; interestingly,

all the adenocarcinomas in the same study appeared genetically homogeneous. More recently, Thirlwell and colleagues [107] performed crypt-by-crypt mutation analysis on both FAP and sporadic human adenomas, and were able to identify multiple genetically distinct sub-clones within the same lesion.

There are two possible explanations for the polyclonality of adenomas observed in the above studies: (1) due to there being a high density of adenomas in both the FAP mucosa and the mouse model, *random collision* of neighbouring lesions may occur; (2) In order for a dysplastic clone to escape the inhibitory effects of surrounding normal tissue, such as BMP signalling [61], and grow, *clonal interaction* is required for adenoma formation [108]. To assess the contribution of random collision, homozygosity for the tumour resistance allele of the *Mom1* locus was introduced into *ROSA26*/chimeric Min mice in order to allow analysis of clonality at a much lower tumour burden [109]. The results showed that the proportion of polyclonal adenomas was much greater than that predicted by a random collision hypothesis: examination of clonal interaction using spatial statistics indicated that co-operation occurred between clones as close as 1-2 crypts apart via short-range interactions. Evidence has shown oncogenic genetic alterations occurring in surrounding stromal tissue of colorectal adenomas at an early stage of carcinogenesis [110], and the epithelium immediately adjacent to colorectal cancers is known to be hyperplastic [17]. In addition, concordant mutation of TGFBR1*6A – a common hypomorphic variant of the type I transforming growth factor (TGF)-beta receptor (TGFBR1) that confers a selective growth advantage to normal and cancerous epithelial cells as they are less susceptible to the inhibitory effects of TGF-β, and is somatically acquired in colorectal cancer cells - has been demonstrated in the stroma and epithelium adjacent to colorectal neoplasms [111]. Thus, a dysplastic clone may induce changes in the surrounding stroma and microenvironment that in turn transforms neighbouring crypts, allowing expansion of the lesion to form an adenoma, with subsequent progression to a carcinoma (Figure 1.5).

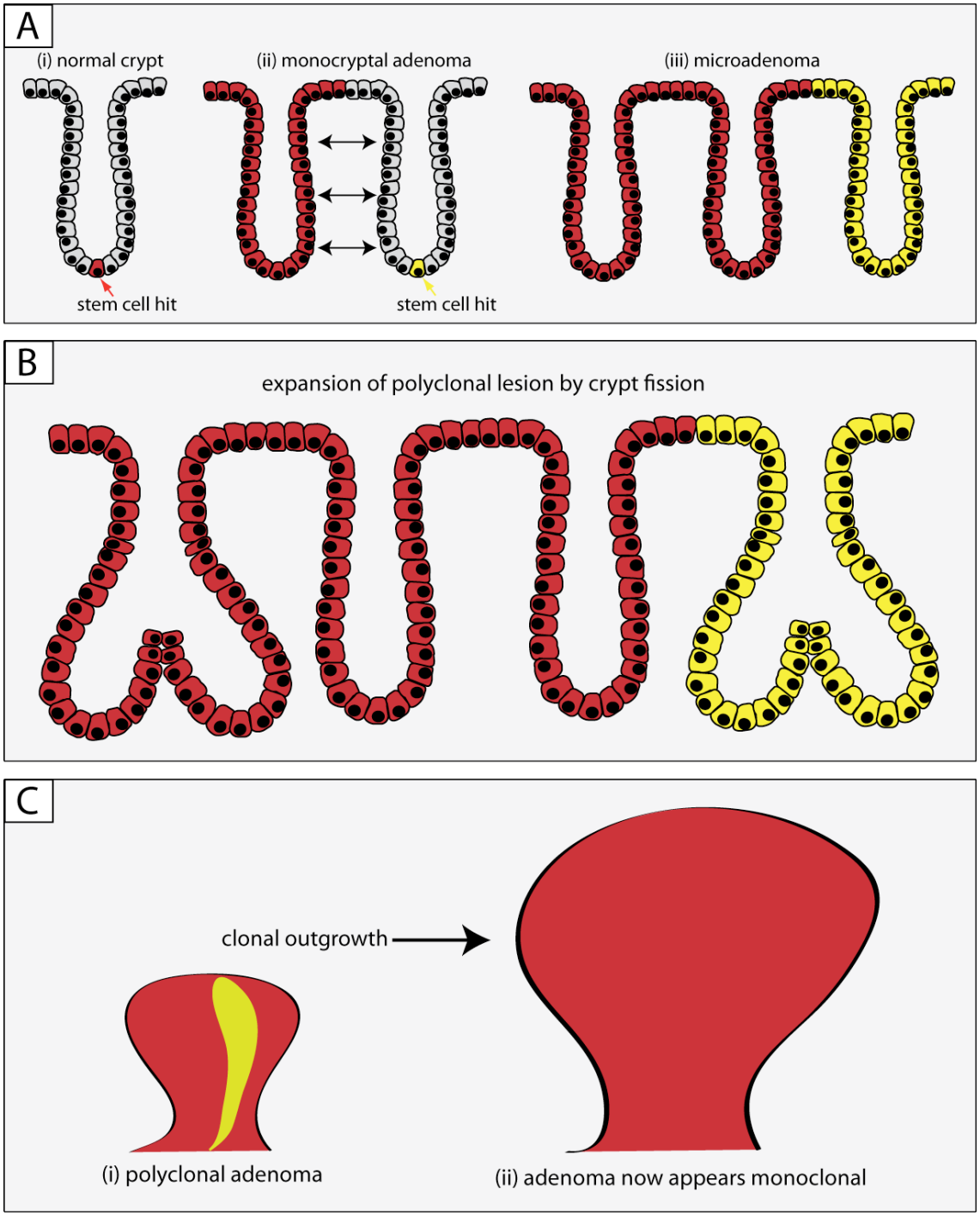

Figure 1.5 Polyclonal adenoma growth
(A): A stem cell develops a mutation (i), possibly in the *APC* gene, and colonizes a crypt with *APC*-/- cells to give a monocryptal adenoma (ii). In some way, possibly through altered signalling within the lamina propria, mutations are induced in an adjacent crypt(s) – clonal interaction, which also clonally-converts, forming a polyclonal lesion (iii). The adenoma then expands by crypt fission **(B)**, until eventually, expansion of a dominant clone leads to a more advanced adenoma that appears monoclonal **(C)**.

Therefore, the mechanisms described above detail how a single dysplastic colon crypt can arise and expand via crypt fission to become an established adenoma. During adenoma growth the normal crypt hierarchy appears to be maintained [112], but significantly increased rates of crypt fission are occurring and there is an upward expansion of the proliferative compartment towards the surface of the crypt [94], consistent with deregulation of the crypt cycle, to enable rapid growth. Further mutations in *KRAS* and subsequently other critical tumour suppressor genes, such as Tumour Protein 53 (*TP53*), are temporally associated with the progression of the adenoma to a carcinoma [113, 114] (Figure 1.6). Detailed studies in human adenomas have suggested that that they can be polyclonal – distinct *APC* and *KRAS* sub-clones can be identified in adenomas [106, 107], subsequently a selective sweep may occur whereby the successful sub-clone sweeps to dominance and the lesion appears monoclonal once more (Figure 1.5C).

1.3 Should we always blame the stem cell?

It is thought that, in renewing systems, the origins of cancer are in the stem cell [114], as they have been thought to be the only cells that live long enough to acquire the necessary multiple mutations required for neoplastic transformation, moreover they already possess some of the characteristics of cancer cells, such as self-renewal [115]. The first direct supporting experimental evidence that transformation of stem cells is the principal route to intestinal cancer was recently provided by the Clevers' group [116]: deletion of *Apc* limited to the Lgr5+ stem cells of the mouse intestine resulted in the rapid formation of microadenomas that expanded to large numbers of intestinal macro-adenomas; however when *Apc* was deleted in the transit-amplifying cells, adenomas were significantly fewer and only able to reach a very small size. Interestingly, they also noted that the crypt stem/progenitor cell hierarchy was maintained in the early neoplastic lesions that developed after mutating *Apc* in the stem cell compartment.

However, there is definitive evidence for the existence of long-lived committed progenitor cells in the mouse intestine [17] and stomach [117] – cells that are committed to the production of a limited repertoire of descendants. In the intestine, long-lived mucous cell progenitors and columnar cell progenitors have been found, which in the mouse have a life-span measured in months [17]. Similarly, in the gastric epithelium, after chemical mutagenesis which labels random epithelial cells by loss of transgene function in adult hemizygous *ROSA26* mice, clones containing only a single mature cell type appear, indicating the presence of long-lived committed progenitors in the gastric epithelium. The same phenomenon has been demonstrated in the human stomach for the parietal cell series - the only wild-type cell lineage present in occasional CCO-mutated glands were parietal cells, strongly suggesting that the human stomach contains long-lived committed progenitors [78]. This raises the possibility that the cell that undergoes the first mutation and selection event is not always the stem cell, but a committed progenitor cell [118]. While it has to be said that there is no direct evidence for this, particularly in the human model, the possibility that this occurs would explain several common observations: tumors arise from the colon which are composed predominantly, or exclusively, of mucin-secreting cells [119], endocrine cells [120] or even Paneth cells [121]. Thus, it may be that it is not only the stem cell that has the potential to initiate the sequence of successive mutations and selection demanded by the concept of multistage carcinogenesis [122].

1.4 Genetic changes and tumorigenesis

The *adenoma-carcinoma* sequence is widely accepted as the pathway for evolution of cancer in the colon, and has been well defined; it is a classical example of tumorigenesis involving the successive accumulation of genetic changes that accompany histological progression through dysplasia to adenocarcinoma (Figure 1.6). The study of the genetic basis to the two main hereditary syndromes associated with the development of colorectal cancer, FAP and Hereditary Non-polyposis Colorectal Cancer (HNPCC), have also given many insights into the genetic changes associated with sporadic colonic tumorigenesis; in a similar fashion the more recent recognition of the malignant

potential of Hyperplastic Polyposis Syndrome (HPS) has resulted in the identification of alternative and genetic pathways to cancer [123]. As molecular techniques have improved, the original model proposed by Vogelstein [114] is being replaced with a number of genetic pathways that may result in colorectal cancer (Figure 1.6).

1.4.1 Mechanisms of fixation

The processes of niche succession and monoclonal conversion allow tumorigenic mutations to be fixed in a number of ways. Due to the small number of stem cells within the colonic niche, they are prone to *genetic drift*, therefore a single stem cell clone can come to dominate the niche as a result of stochastic non-selective events. As a result, neutral mutations that do not affect the fitness if a cell can, over time, become fixed within the population [124]. Occasionally, mutations in a stem cell, especially those involving a tumour suppressor gene such as *APC*, bestow a selective growth advantage to the recipient cell. This allows the processes of niche succession and monoclonal conversion to occur rapidly so that the mutated stem cell progeny occupy the whole crypt. These mutations often result in what is termed a *selective sweep* where *natural selection* rapidly drives the advantageous allele to fixation [125-127]. Although tumorigenesis involves the accumulation of mutations that result in a growth advantage to the cell, not all confer an immediate selective advantage in isolation, but in combination with others they contribute to a malignant phenotype. These mutations may become established merely by being linked to or *hitchhiking* with an allele that does provide a selective advantage [50, 128].

1.4.2 Sporadic adenomas: A multi-step process

Patients with FAP have a germ-line mutation in the tumour suppressor *APC* gene and this has been termed the *gatekeeper* of colonic epithelial cell proliferation [129], as once both alleles are lost in FAP patients many thousands of polyps are able to form but the further genetic changes required for the progression to carcinoma occur at the normal

rate. Due to the huge number of adenomas it is inevitable that one of these will eventually undergo malignant transformation, hence FAP patients are offered a prophylactic colectomy early in life. In HNPCC, or Lynch syndrome, the mechanisms are different: Germ-line mutations in the DNA mismatch repair genes (MMR), commonly *MLH1, MSH2* and *MSH6,* are linked to microsatellite instability (MSI) and result in higher mutation rates of repetitive sequences [130]. The incidence of adenoma formation in these patients is the same as the normal population but there is accelerated tumour progression: MMR genes have thus been termed *housekeeping* genes.

The hereditary syndromes FAP and HNPCC themselves only account for less than 5% of colorectal cancers [131], but the application of knowledge gained from understanding the genetics involved has enabled elucidation of some of the major molecular processes involved in the evolution of sporadic carcinomas. *APC* mutations are found in up to 80% of sporadic colorectal adenomas and cancers [84, 132, 133] and this is thought to be the initiating step in the development of most sporadic adenomas; the remaining 20% may contain as yet undetected *APC* or *β-catenin* mutations. Deficiency of MMR repair (MSI) is found in approximately 15% of all colorectal cancers, mostly via acquired hyper-methylation of the *MLHI* promoter [99, 134, 135], but less than 3% of sporadic adenomas; therefore, in sporadic cases, MSI is thought to occur at a more advanced stage in adenoma progression. As only a very small population of the thousands of adenomatous polyps that form in the colon of FAP patients progress to cancer, it is obvious that further oncogenic mutations are required for a tumour to develop. Molecular studies of colorectal cancers have confirmed that multiple sequential genetic mutations are required for tumour formation, these can be thought of as independent events, but the higher intrinsic mutation rates found in those tumours with defects in MMR genes, and the gross chromosomal alterations seen within sporadic tumours with no MSI results in increased genetic instability and predisposition to subsequent mutational events [129].

The KRAS protein mediates the intracellular transduction of signals from a number of growth factor receptors, such as epidermal growth factor receptor (EGFR) and platelet-derived growth factor receptor (PDGFR), predominantly via the mitogen-activated protein kinase (MAPK) cascade. Activating *KRAS* mutations, specifically in codons 12 and 13 of exon 2, are detected in approximately 50% of sporadic colorectal adenomas and carcinomas [113, 136, 137], and are an important, frequent event in the development of early sporadic adenomas.

1.4.3 Mutator phenotype or selection

Mutation and selection are the essential components in the evolution of cancer. However, the relative importance of each to tumorigenesis has been the subject of much debate [122, 138, 139]. Mutations in normal cells are rare [139], and it has been suggested that the development of a *mutator phenotype* – cells that have intrinsically elevated mutation rates due to mutations in genes responsible for DNA fidelity, is required early on in tumorigenesis to explain the accumulation of mutations required for multistage carcinogenesis [139, 140].

However, the evidence as to when the mutation rate can become elevated in colorectal cancer is not clear [138], and it has been shown that the selection of mutations that confer a growth advantage to a cell can enable successive rounds of clonal expansion, or selective sweeps [125-127]. Knudson [141], studying hereditary and non-hereditary retinoblastoma, proposed what became known as the *two-hit hypothesis*, whereby the initial rate-limiting step in cancer is the loss of two alleles of a tumour suppressor gene, one of which is inherited and the other selected for, and this is sufficient to initiate clonal expansion, thus a disease with a recessive inheritance could have a dominant penetrance if there was a high likelihood of selection for the loss of the second allele. Tomlinson et al [138] demonstrated that selection, more than an increase in the intrinsic mutation rates, is more likely to be the driving force in sporadic tumorigenesis, and

Luebeck et al [142] argued that genomic instability is not required to explain the clonal expansion of an initiated stem cell into an adenoma. The epidemiology of FAP supports a low mutation rate: despite FAP patients having a germ-line mutation in *APC*, only a tiny fraction of epithelial stem cells become neoplastic [129]. Moreover, studying the associations of the locations of germ-line and somatic *APC* mutations in adenomas from FAP patients, Lamlum et al [143] showed that mutations in codons 1296 or 1309 are strongly associated with allelic loss as the second somatic hit, providing evidence that *APC* mutations offer differing selective advantages to a cell that results in a selective pressure for a specific mutation in the remaining allele.

Thus, the initiation and expansion of a tumorigenic stem cell clone into an adenoma does not necessarily require the presence of a mutator phenotype [122, 138, 142]. However, it may be that as successive rounds of selective sweeps and clonal expansion occur then genomic instability increases, either thorough defects in MMR genes or gross chromosomal loss, so that both selection and an elevated mutation rate drive further evolution to cancer.

1.4.4 Traditional Pathway

The seminal work by Vogelstein and colleagues [113, 114, 144] first formulated that colorectal cancer was the result of a stepwise accumulation of genetic alterations in at least 4 or 5 genes, combining mutational activation of oncogenes and inactivation of tumour suppressor genes. They classified adenomas as being early, intermediate or late on the basis of overall size and histology: early adenomas were those less than 1cm; intermediate adenomas were those greater than 1cm and with no foci of carcinoma; late adenomas were those that were greater than 1cm and contained foci of carcinoma. The frequency of *RAS* mutations and chromosomal allelic loss of 5q (*APC)*, 17p (*TP53)* and 18q detected in the adenomas and carcinomas, combined with a further study looking at *TP53* mutations associated with allelic loss of 17p [144], led them to propose a sequence

of genetic events that occurs during colorectal tumorigenesis [113, 114]: *APC* loss is the gate-keeping mutation, *KRAS* mutations develop in intermediate adenomas with *TP53* mutation, 17p and 18q loss of heterozygosity (LOH) occurring in advanced adenomas as they progress to cancer (Figure 1.6A). The *SMAD4* gene, a member of the transforming growth factor-β (TGF-β) pathway, is now regarded as the most likely tumour suppressor involved with 18q loss [145].

It is worth noting, and this is something that is often ignored, that the model proposed by Vogelstein summarised only the most frequent genetic events detected: 17p loss was still found in almost 10% of intermediate adenomas and a few early adenomas; regions from chromosomes lq, 4p, 6p, 6q, 8p, 9q, and 22q were lost in 25%~50% of the colorectal cancer cases studied. Screening was done on DNA digested from entire lesions, thus any small sub-clones containing different mutations would not have been identified. Also, these studies only looked at conventional adenomas i.e. tubular and tubulo-villous, as the malignant potential of hyperplastic and serrated lesions had yet to be recognised.

1.4.5 Alternative and serrated pathways

The original model proposed by Vogelstein and colleagues [114] has since been expanded, with the recognition of alternative genetic pathways to cancer. There is now acceptance that activation of the mitogen activated protein kinase pathways, via mutations in either the *KRAS* or *BRAF* oncogenes, combined with acquired MSI and methylation of promoter CpG sites in a number of genes (CpG island methylator phenotype – CIMP), are important pathways in the progression to colorectal cancer that are independent of *APC* mutation [99, 123] (Figure 1.6). CIMP refers to specific, age independent, methylation of a number of CpG islands in the promoter regions of a panel of genes implicated in carcinogenesis (*CDNK2A (p16), MLH1, MINT1, MINT2, MINT31, RUNX3, CACNA1G, IGF2, NEUROG1, SOCS1*) [135, 146]; lesions are often classified as either CIMP-negative,

CIMP-low if one or two CpG islands are methylated or CIMP-high if three or more of the panel are methylated. A distinct subset of colorectal cancers display CIMP-high, and they tend to be proximally located, poorly differentiated, *TP53* wild-type and are highly associated with *BRAF* mutations and MSI [135, 146, 147].

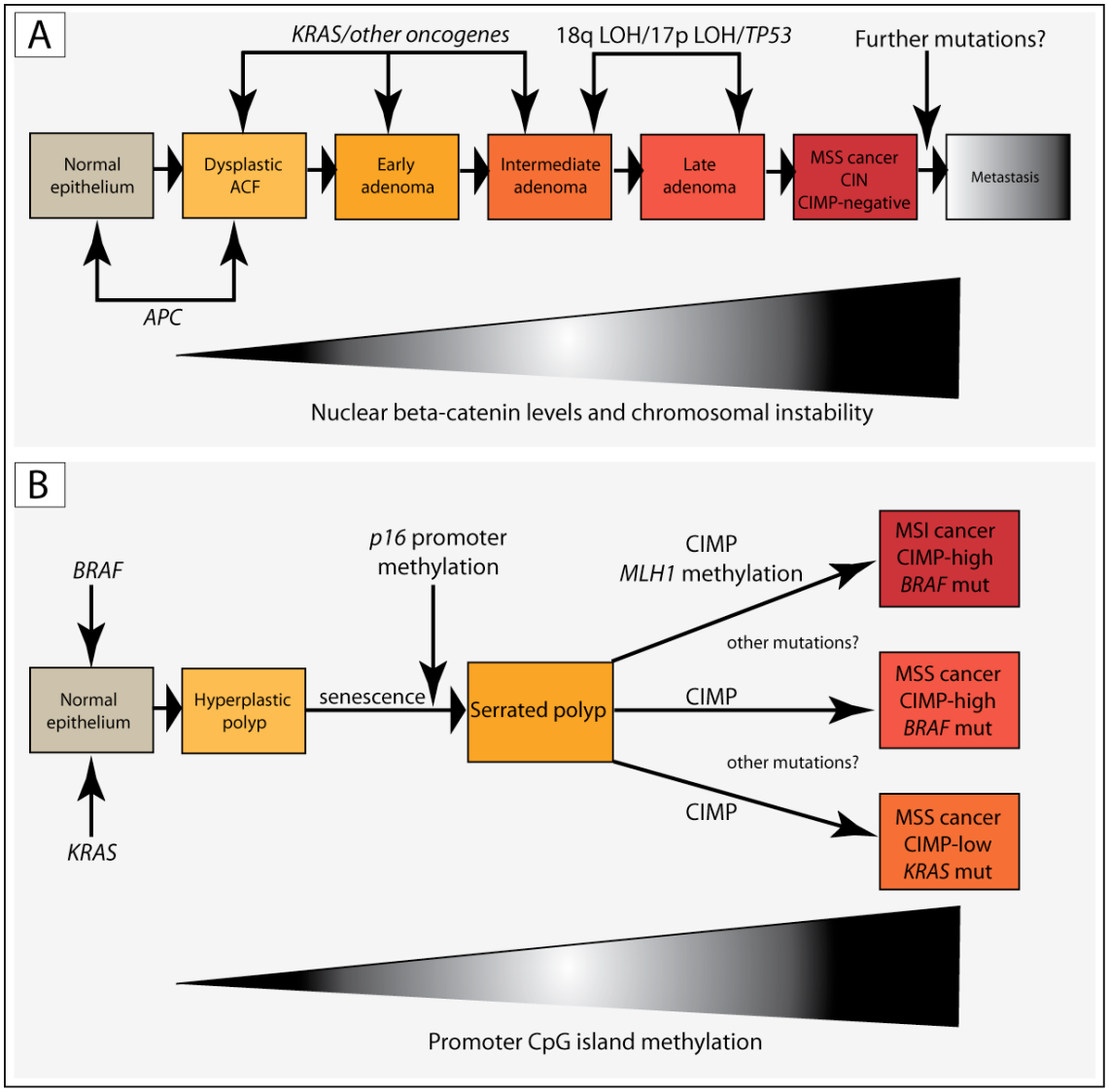

Figure 1.6 Traditional and alternative genetic pathways in colorectal tumorigenesis
(A) Traditional pathway: mutations in *APC* are required for the initiation of most sporadic adenomas, with *KRAS* mutations also occurring early on in adenoma development; loss of 18q and 17p in addition to mutations in *TP53* are thought to be required for the progression to cancer. Most sporadic cancers are MSS, do not display CIMP but have a high level of CIN. **(B) Alternative and serrated pathways:** Mutations in either *BRAF* or *KRAS* enable a proliferative burst that is followed by *p16* induced senescence; the development of CIMP allows escape and progression to serrated polyps; *BRAF* mutations are associated with a high level of CIMP and MSI cancers, whereas *KRAS* mutant lesion tend to be CIMP-low and MSS. This is a simplification of the predominant pathways, and there is overlap between the two, with many other oncogenes and tumour suppressor genes likely involved that have yet to be fully elucidated. Abbreviations: MSS microsatellite stable; MSI microsatellite instability; CIN chromosomal instability; CIMP CpG island methylator phenotype.

The common feature of serrated polyps is a *saw-toothed* infolding of the epithelium. Previously these were classed with hyperplastic polyps and thought to have no malignant potential, however, a clear association between serrated polyps, especially sessile serrated adenomas, and sporadic MSI colorectal cancer has now been established, with cancers demonstrated to have directly arisen from adjacent serrated polyps with a transition zone of dysplasia [148]. The *serrated pathway* is thought to be initiated by mutations in either *BRAF* or *KRAS*, the two being mutually exclusive although *BRAF* mutations are more common [149], this is followed by *p16* induced senescence which is overcome by methylation with the early development of CIMP; MSI is then associated with later progression of the lesions [123, 149] (Figure 1.6B). A single point mutation (V600E) accounts for nearly all *BRAF* mutations, resulting in constitutive activation of the MAPK pathway downstream of KRAS. Interestingly, although *BRAF* mutations are very common in serrated polyps, they are almost never seen in conventional adenomas [150]. However, there now appears to be a clear, tight association between *BRAF* mutations and the development of CIMP and MSI in colorectal cancers [135, 150].

In summary, there are multiple genetic pathways involved in colorectal tumorigenesis; mutations in *APC a*nd *KRAS* still appear to be the commonest genetic events in early adenomas, however alternative pathways, specific to the histology and location of the lesion in the colon, exist and are being increasingly recognised. Fearon and Vogelstein concluded that '*it is the accumulation of events, rather than their order of occurrence with respect to one another, that is likely to be most important in colorectal tumor progression*' [114], however it is now thought that the timing of genetic events is crucial in determining whether a lesion will be able to progress to cancer [123, 151].

1.5 Rates of adenoma growth

Although adenomas are the precursor lesion to colorectal adenocarcinoma, they have a low malignant potential, with previous studies suggesting it takes in the order of 15 years for a large adenoma to become malignant [152], consistent with the 20-25 years it takes for small adenomas to grow into adenocarcinomas in FAP patients. Stryker et al [153] demonstrated that the cumulative risk of cancer for large adenomas (>1cm) at 5, 10 and 20 years was only 2.5%, 8% and 24% respectively. A more recent study [154] used large-scale sequencing to compare mutations detected in colorectal metastasis, adenocarcinomas and adenomas of the same patient – so termed *comparative sequencing*. By quantifying the numbers of mutations acquired as the lesions progressed, and combining this with the clinical time points, the authors estimated that it may take, on average, 17 years for a benign sporadic adenoma to progress to cancer, assuming that mutation rates and cell division times do not change throughout the patient's life.

Early estimates of adenoma growth rates were based on the analysis of size doubling times of adenomas using retrospective serial barium enema studies [155-157]. Welin et al [155] reviewed over 21,000 barium studies over a period of 10 years. They observed that adenomas <10mm appeared to grow very slowly, if at all, and, by mathematically modelling growth data, concluded that 10mm is the critical size threshold for predicting future growth. Of 11 adenomas studied by Tada et al [157], 5 did not grow at all during the follow-up period, which ranged from 420 to 816 days; doubling times of the other 6 adenomatous polyps ranged from 146.5 to 398.5 days.

More recent endoscopic studies have supported the original findings by Welin et al [155] that a size of 10mm appears to be critical [158-160]. Hofstad et al [159] conducted a prospective colonoscopic study where adenomas <10mm in size where left in-situ for 3 years and those >10mm were resected, polyp size was accurately measured using a

probe inserted through the endoscope and all polyps >5mm were removed at the end of the study. Interestingly, the authors found that, for adenomas and hyerplastic polyps, lesions <5mm showed a tendency to growth but medium sized polyps of 5-9mm demonstrated a net reduction in size. Bersentes et al [160] followed thirty polyps of 3-9mm endoscopically over two years: they did not find any regression, however only 3 adenomas showed fast growth rates (2-4mm/year) overall and growth did not follow a linear pattern. In an observational study by Knoernschild [161] 18% of 257 polyps that were tattooed and followed up for 3-5 years had completely disappeared.

In summary, it appears that adenomas may not grow in a constant linear manner and appear to remain relatively static for many years, with a large proportion of smaller lesions even regressing over time, which is consistent with observations that adenoma incidence does not increase with age [162, 163].

Chapter 2. Materials and Methods

General methods used are detailed here. Any variation from the techniques detailed below is discussed in the appropriate results chapter. Specific methods that were designed and developed as part of a project are also given below, and then referenced in the respective results chapters.

2.1 Tissue collection and sectioning

2.1.1 Fresh frozen tissue

Normal human colon and polyps were obtained from patients undergoing surgery or colonoscopy screening at University College Hospital, London. Ethical approval was obtained as per the requirements of the United Kingdom Human Tissue Act (2006), REC reference number 07/Q1604/17. Fresh tissue was immediately frozen in liquid nitrogen using isopentane as a cryoprotectant, and then stored at -80°C until further use.

Serial sections of tissue were then cut either ‘on-edge’ or ‘en-face’ depending on the project requirements (detailed later), at thicknesses varying from 8 microns to 20 microns. If being prepared for laser-capture micro-dissection (LCMD) then sections were mounted onto P.A.L.M. laser capture slides (P.A.L.M. Microlaser, Technologies, Benried, Germany) that had been pre-treated with ultraviolet radiation, at a wavelength of 254 nanometres (nm), to improve section adhesion. For tissue intended for immunohistochemistry, sections were cut at 6-8 microns and mounted onto charged glass slides (Thermo-Scientific, Germany).

2.1.2 Formalin-fixed paraffin embedded tissue

Archived normal human colon blocks were obtained from University College Hospital London in accordance with the requirements of the United Kingdom Human Tissue Act (2006), REC reference number 07/Q1604/17. Serial sections were cut at 6 microns onto charged glass slides (Thermo-Scientific, Germany).

2.2 DNA extraction

Proteinase K digestion buffer (Arcturus Bioscience, USA) was used for all DNA extraction. All equipment, gloves and reagents, excluding proteinase K enzyme, were sterilised by 45 minutes of ultraviolet radiation at a wavelength of 254 nanometres (Ultraviolet hood, ASTEC Microflow, UK). Prepared proteinase K digestion buffer was kept on ice to prevent auto-degradation.

2.2.1 Tissue macro-dissection

To enable somatic mutation and loss of heterozygosity (LOH) screening of genes commonly involved in early adenoma initiation and growth, and global genomic DNA methylation analysis, 3 serial sections from the middle of each polyp were needle macro-dissected into a 1.5 millilitre micro-centrifuge tube containing 100 microlitres of Proteinase K digestion buffer. Tubes were incubated for a minimum of 16 hours at 65°C, briefly centrifuged and then incubated for 10 minutes at 95°C to deactivate proteinase K. Tubes were then centrifuged at maximum speed for 30 minutes and clear lysate removed via pipette into a clean 1.5 millilitre micro-centrifuge tube, in order to remove residual protein and inert membrane. DNA concentration was then measured using a NanoDrop® analyser and lysate diluted to a DNA concentration of 100-200 nanograms per microlitre as this is optimum for PCR. DNA lysate was then stored at -20°C until further use.

2.2.2 Laser capture micro-dissection (LCMD)

LCMD of single crypts or single cells from normal tissue, adenomas and polyps was performed using the P.A.L.M. Laser Micro-dissection system (P.A.L.M., Zeiss, Germany). Crypts or single cells were micro-dissected and then catapulted into the adhesive caps of the P.A.L.M. laser capture micro-dissection tubes (P.A.L.M., Zeiss, Germany). An appropriate volume of freshly prepared Proteinase K buffer was then added and the tubes incubated at 65˚C for a minimum of 3 hours for frozen tissue and 16 hours for paraffin tissue. Tubes were then centrifuged at 1000x g for 1 minute before heating at 95˚C for 10 minutes to deactivate Proteinase K. Following further centrifugation at 1000x g for 1 minute, DNA lysate was stored at -20˚C until further use.

2.3 DNA amplification and sequencing techniques

2.3.1 Polymerase Chain Reaction (PCR)

Target DNA was amplified using specific primers by PCR. Prior to each reaction, all reagents (excluding primers, and DNA lysate) and sequencing apparatus (including PCR plates, lids, tips and pipettes) were sterilised by 45 minutes of ultraviolet radiation at a wavelength of 254 nanometres (Ultraviolet hood, ASTEC Microflow, UK); first-round PCR was then prepared in the sterilized ultraviolet hood. Reaction mastermixes were prepared and aliquoted into either a 96 or 48-well plate (ABgene, UK). To confirm that the PCR reactions were not contaminated, an additional two negative controls of master-mix only and DNA digestion buffer only were performed. Plates were sealed with adhesive PCR film (Thermo Fisher Scientific, UK) and immediately transferred to a thermal cycler. Where appropriate, second-round PCR was performed on ice. Reaction efficiency was tested by agarose gel electrophoresis (1.5%) of PCR product and successful amplification identified by the presence of a single strong band of

appropriate length: only those reactions where the negative controls were blank on the agarose gel were taken forward for sequencing. See Appendix 9.1 and 9.2 for details of all reagents, primers, PCR and thermal cycler protocols used.

2.3.2 DNA purification and sequencing

PCR products were purified using the ExoSap-IT® (GE healthcare, UK) clean-up kit for enzymatic removal of unused primers and nucleotides; ExoSAP-IT® contains exonuclease I which removes unincorporated primers, and shrimp alkaline phosphatase which removes leftover dNTPs. 2 microlitres of ExoSap-IT® was added to 5 microlitres of each PCR product and then run on a thermal cycler using the protocol detailed in Appendix 9.2 (Table 9.22). All sequencing was performed using the BigDye Terminator cycle sequencing method on an ABI Prism Genetic Analyzer: automated sequencing that uses four terminator nucleotides labelled with distinguishable fluorophores. ExoSAP-IT products were diluted in 10-20 microlitres of distilled H_20, depending on the strength of the PCR product bands on the agarose gel. The sequencing reaction was prepared according to the reagent protocol detailed in Appendix 9.2 (Table 9.20) using the ABI Biosystems BigDye Terminator (Applied Biosystems, USA). For sequencing thermal cycler protocols see Appendix 9.2 (Table 9.23 and Table 9.24). Purified sequencing products were vacuum dried using an Eppendorf 5301 Concentrator (Eppendorf, UK) for 30 minutes at 45°C. Re-suspended residues were transferred to a 48 capillary ABI 3730 DNA Analyser (Applied Biosystems, USA) for sequencing and sequence data analysed with the 4Peaks software (Mekentosj, Netherlands).

2.4 Mitochondrial DNA mutation detection

2.4.1 Enzyme histochemistry

Two-colour enzyme histochemistry was used to simultaneously detect the mtDNA-encoded enzyme cytochrome *c* oxidase (CCO) and nuclear DNA-encoded succinate dehydrogenase (SDH), a component of complex II of the respiratory chain. Mutated cells lacking in CCO-activity appear blue; non-mutated cells appear brown. Frozen sections were air-dried for 45 minutes and first incubated in cytochrome c oxidase medium (100 millimolar cytochrome c, 20 grams per millilitre catalase and 4 millimolar diaminobenzidine (DAB) tetrahydrochloride in 0.2 millimolar phosphate buffer, pH 7.0, all sourced from Sigma-Aldrich, Poole, UK) for a maximum of 50 minutes at 37°C. Crypts or cell populations within crypts with normal activity of CCO subsequently appear brown. In order to counter-stain CCO-deficient cells, sections then underwent 3x 5-minute washes in phosphate-buffered saline (PBS), pH 7.4, before being incubated in SDH medium (130 millimolar sodium succinate, 200 millimolar phenazine methosulfate, 1 millimolar sodium azide, and 1.5 millimolar nitroblue tetrazolium in 0.2 millimolar phosphate buffer, pH 7.0, all sourced from Sigma-Aldrich, Poole, UK) for a maximum of 45 minutes at 37°C, or until a strong blue stain had developed in the unstained crypt cell populations. Sections again were washed in PBS (5 minutes x3) and dehydrated in a graded ethanol series (70%, 90%, 100%, 100%). If tissue was for LCMD slides were air-dried for 1 hour before being stored at -80°C; sections that were on normal glass slides were cleared in Histoclear (Lamb Laboratory Supplies, Eastbourne, UK) and mounted with Permount® (Fisher Scientific, FairLawn, NJ).

2.4.2 Laser-capture micro-dissection

After enzyme staining for CCO-activity, three different types of intestinal crypts are observed: (i) wild-type brown crypts, (ii) wholly-mutated blue crypts and (iii) crypts containing populations of both brown and blue cells (mixed or partially-mutated crypts). LCMD was then performed from both wholly-mutated and partially-mutated blue crypts, and adjacent brown wild-type crypts, in order to sequence the mitochondrial DNA and confirm the presence of clonal point mutations in the blue cells. Laser-capture slides were allowed to thaw at room temperature for 1 hour before micro-dissection. Single cells from CCO-deficient crypts and adjacent wild-type crypts were individually micro-dissected. If required for genomic and/or methylation analysis, the remaining regions of the whole crypt were then dissected from serial sections. 14 microlitres of Proteinase K digestion buffer (Arcturus Bioscience, USA) was added to each tube for a minimum of 3 hours digestion. A negative control tube containing only digestion buffer and no laser-capture material was also included. In partially CCO-deficient crypts, CCO-deficient cells were micro-dissected separately from CCO-proficient cells, and then the remaining blue and brown cell populations from the whole crypt micro-dissected into two tubes for methylation analysis. After digestion, tubes were briefly centrifuged and proteinase K deactivated by heating to 95˚C for 10 minutes. DNA lysate was stored at -20˚C until further use.

2.4.3 MtDNA sequencing

Laser micro-dissection followed by PCR coupled with mtDNA sequencing confirms the presence of clonal mitochondrial mutations in wholly or partially CCO-deficient crypts, and this has been shown to be a reliable marker for observing the clonal expansion of mutated stem cells [76-78]. Extracted DNA lysate was used to sequence the entire mitochondrial genome from micro-dissected areas. A two-round amplification method was followed, whereby the first round consisted of amplifying 9 fragments spanning the entire mitochondrial genome. These initial large PCR products of approximately 2kb

decrease the risk of amplifying pseudo-genes when using DNA extracted from small quantities [164]. The second round consisted of 36 M13-tailed primer pairs to amplify overlapping segments of the first-round products, resulting in amplicons of 500-700 base-pairs that span the entire mitochondrial genome (see Appendix 9.1 and 9.2 for primer sequences and detailed PCR protocols). PCR product from the second round reaction was run on a 1.5% agarose gel, and only those reactions where the negative controls were blank on the agarose gel were taken forward for sequencing. DNA purification and sequencing was performed as previously described (see Methods 2.3.2) and then compared to the revised Cambridge reference sequence (http://www.mitomap.org/MITOMAP) using sequence alignment software of the European Bioinformatics Institute (http://www.ebi.ac.uk). Once a potential mutation was identified, the PCR was repeated using the original DNA lysate and multiple samples from the mutated blue cells and adjacent wild-type, brown cells in order to confirm its validity.

2.5 Mapping of stem cell clones

By combining the two-colour enzyme histochemical staining with image analysis and computer reconstruction, a method of mapping the origin and spread of stem cell progeny and mutated cell populations within the intestinal crypt was devised. Multiple serial sections in the transverse plain were taken through frozen samples at thicknesses between 8 and 12 microns, depending on the quality of the tissue. Digital images were then taken of each serial section and used to create a 'crypt map' using image analytical algorithms and in-house software developed by Dr Paul J Tadrous (Histopathology Dept, Northwick Park Hospital, Harrow, London). A crypt map is a representation of the whole 3D tubular crypt unfurled and laid flat – like a map – with colour enhancement post-processing designed to increase the contrast between blue-staining and non-blue-staining cells. The algorithm for producing such a crypt map is as follows:

1. Serial section images of the crypt are registered (aligned) to form a 3D reconstruction series of the whole crypt.

2. The centre-point and perimeter of cross sectional image of the crypt in such a series is then marked out by hand using freely available interactive image analysis software (BiaQIm v 2.5 alpha from www.bialith.com).
3. For each cross sectional image a line is then subtended from the centre-point to the perimeter at a fixed angle (known as the 'cut-angle'). The cut-angle is the same for all cross sectional images in a given crypt series.
4. For each image in the series, the approximately oval cross sectional profile of the crypt is then digitally 'cut' along the line described in step 3 and is warped into a straight profile.
5. For each straight profile thus generated, the basal-most pixels (i.e. the pixels starting at the basement membrane aspect and continuing towards the centre for a distance of about ¼ of the epithelial thickness) are then averaged into a single pixel strip. This step acquired most of the cytoplasmic colour information before the cytoplasm gets 'diluted' by mucin as one gets close to the lumenal aspect of the crypt epithelium.
6. These single pixel-strips are then arranged in order of the 3D series – from crypt base at the bottom to gut lumenal surface at the top. This forms the basic 'crypt map' without colour enhancement.
7. The colour post-processing of this basic crypt map is as follows (RGB = red, green, blue; HSV = hue, saturation, value i.e. intensity value; LUT = look-up table):
    7.1 Transform the 24 bit RGB components into HSV channels
    7.2 Replace the V channel of this transform with a scaled version of the S channel (scaled such that the values of the S channel lie in the range of 0 to 255)
    7.3 Recombine these modified HSV channels and back-transform to give an RGB colour image where the intensity of each pixel is now directly proportional to the colour saturation of the original image.
    7.4 Select the B channel of this new image and display the intensity values on a logarithmic intensity scale to give greater visibility to subtle blue shades. This final image is a monochrome image but is displayed in a blue pseudo-colour LUT for display purposes.

The end result of the above is a crypt map where the intensity value of each pixel is a log-transformed representation of the blue component in the original RGB image multiplied by its saturation. Thus a pixel that is white in the original image (e.g. mucin or other unstained areas) will have lots of blue in it but will be very unsaturated so will be dark in this image. A pixel which is brownish, reddish or greenish in the original image will have a high saturation value but will not have a hue in the blue range so will appear dark in this enhanced image. Only those pixels that have a *both* a hue in the blue range *and* are also saturated will appear bright in this post-processed image. Thus these final crypt maps display a representation of visual blue staining in the crypt.

By this method the mutated (blue) clonal cell lineage is differentiated from the non-mutated (brown) cells and any non-staining areas – the latter appearing black on the final crypt maps. The crypt map thus shows the path taken by the mutated blue clone that can be fully visualised from crypt base to crypt lumenal surface.

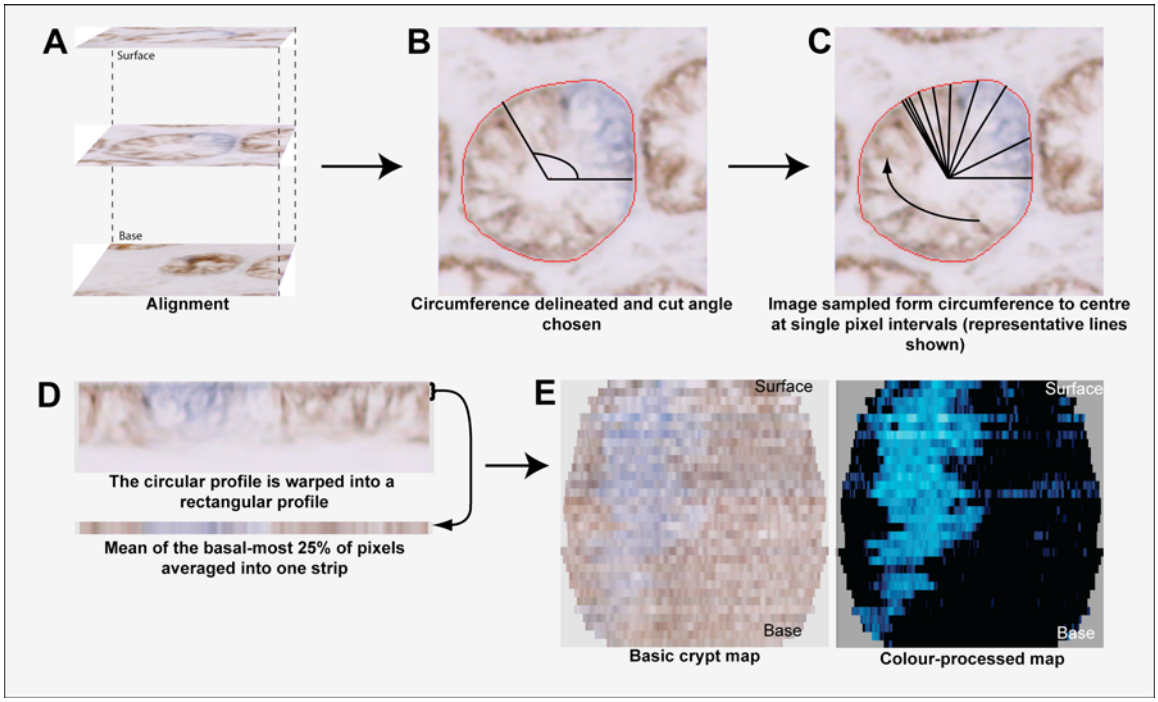

Figure 2.1 Construction of the crypt map
Multiple serial sections are cut through frozen epithelial tissue and then stained for CCO-activity; partially-mutated crypts are identified and serial sections photographed. The images series is then aligned to a sub-pixel level (A). For each image, the circumference of the crypt is delineated, the centre of the crypt lumen marked and a cut angle chosen (B), the image is then sampled along a spoke radiating from the circumference to the centre at single pixel intervals (C). Each sampled spoke then becomes one vertical line in a new rectangular image (D), where the basement membrane (BM) is along the top edge. A section is taken from the BM aspect and averaged onto one plane which the forms one scan line of the basic crypt map, this is then repeated for each image and the scan lines stacked on top of each other to form the basic crypt map (E), all non-blue cells then appear black in the final colour-processed map in order to allow better identification of the blue clone (E).

2.6 Immunohistochemistry

All formalin-fixed paraffin embedded (FFPE) sections were de-waxed to water, with an additional hydrogen peroxide blocking step for endogenous peroxidases if required, and antigen retrieval performed by micro-waving in boiling citrate buffer (pH 6.0) for 10 minutes; sections were then washed in cold water before being transferred to PBS. Frozen sections were thawed at room temperature for 30 minutes before fixing in cooled Acetone (-20°C) for 10 minutes; sections were then left at room temperature for 30 minutes to allow for evapouration of acetone before being transferred to PBS. All antibodies were diluted in PBS.

2.6.1 Chromogen

Sections were first blocked with a serum-free protein block (Dako, UK) for 10 minutes followed by incubating with streptavidin for 15 minutes and then biotin, also for 15 minutes, at room temperature (Vector Laboratories, Peterborough, UK). The primary antibody - mouse anti-human CCO subunit 1 IgM (1:500; Invitrogen, UK) - was applied for 35 minutes at room temperature. Sections were then washed in PBS (5 minutes x3) followed by 35 minutes incubation with a biotinylated anti-mouse immunoglobulin (Dako, UK). After the final PBS washes (5 minutes x3), sections were incubated with streptavidin peroxidase (Dako, UK) for another 35 minutes, washed in PBS (5 minutes x3), and developed in a solution containing 4 millimoles per litre diaminobenzidine and 0.2% hydrogen peroxide (Dako, UK) for 2-5 minutes. Sections were then dehydrated through alcohol (75%, 95% and 100%), cleared with xylene, and mounted with DePeX.

2.6.2 Fluorescence

Sections were blocked for 10 minutes in serum-free protein block (Dako, UK) and then washed in PBS. Incubation with the primary antibody was for 60 minutes, those used

were: rabbit anti-human MUC-2 IgG (1:100 dilution; Abcam, UK); mouse anti-human MUC-2 IgM (1:100 dilution; Vector Laboratories, UK); sheep anti-human Carbonic Anhydrase II (CA-II) IgG (1:100 dilution; Abcam, UK); mouse anti-human CCO subunit 1 IgM (1:400 dilution; Invitrogen, UK); mouse anti-human CD-10 (1:100 dilution; Dako, UK). Sections were then washed in PBS (5 minutes x3). Appropriate Alexa Fluor® (Molecular Probes, Invitrogen, UK) secondary antibodies were then applied (1:150 dilution) for 60 minutes. Sections were again washed in PBS (5 minutes x3) before being cover-slipped using VECTASHIELD® HardSet™ Mounting Medium with DAPI (Vector Laboratories, UK). Sections were examined using an Olympus BX61 epi-fluorescence microscope system with SmartCaptureX software (Digital Scientific, Cambridge, UK) to generate RGB images from multi-channel monochrome captures. Negative controls that lacked the primary antibody were run in tandem. All primary and secondary antibody dilutions were optimised, and detailed protocols for the double and triple immunofluorescence (IF) performed on fresh frozen and FFPE human colon respectively in Chapter 4 are given below:

Double immunofluorescence protocol for MUC-2 and CA-II on fresh frozen normal human colon:

Fresh frozen sections were thawed at room temperature for 30 minutes, fixed in cooled acetone (-20°C) for 10 minutes and then air-dried at room temperature for 30 minutes before being transferred to PBS. Sections were then incubated for 10 minuets in serum-free protein block (Dako, UK) followed by a further 10 minutes in donkey serum (10% diluted in PBS). The primary antibodies mouse anti-human MUC2 and sheep anti-human CA-II (1:100 dilution in PBS) were applied for 60 minuets and sections then washed in PBS (5 minutes x3). The following secondary Alexa Fluor® fluorescent antibodies were then applied for 60 minutes: donkey anti-mouse (green/488nm) and donkey anti-sheep (Cy5/647nm) all diluted to 1:150 in PBS. Sections were then washed in PBS (5 minutes x3), mounted in VECTASHIELD® HardSet™ and left to dry in the fridge (+4°C) overnight.

Triple immunofluorescence protocol for CCO, MUC-2 and CA-II on FFPE normal human colon:

Sections were de-waxed to water with antigen-retrieval performed as previously described, they were then incubated for 10 minutes in serum-free protein block followed by a 10 minute block in donkey serum (10% diluted in PBS). The following primary antibodies were applied for 60 minutes: mouse anti-human CCO subunit 1 IgM at 1:400, rabbit anti human MUC2 IgG at 1:100 and sheep anti-human CA-II IgG at 1:100 (diluted in PBS). Sections were washed in PBS (5 minutes x3) and then incubated with the following secondary Alexa Fluor® fluorescent antibodies for 60 minutes: donkey anti-mouse (green/488nm), donkey anti-rabbit (Cy3/555nm) and donkey anti-sheep (Cy5/647nm) all diluted to 1:150 in PBS. Sections were then washed in PBS (5 minutes x3), mounted in VECTASHIELD® HardSet™ and left to dry in the fridge (+4°C) overnight.

2.7 *In-situ* hybridization for human *LRG5* mRNA

In-situ hybridisation (ISH) for human *LGR5* mRNA was performed on FFPE normal human colon, all experimental procedures were performed with the kind assistance and close supervision of Mrs Rosemary Jeffery (Histopathology Lab, London Research Institute, Cancer Research UK, London, UK).

2.7.1 Human *LGR5* ISH riboprobe design and synthesis

An ISH probe that was specific to human *LGR5* mRNA was kindly provided by Dr Stefania Segditsas (The Molecular and Population Genetics Lab, Wellcome Trust Centre for Human Genetics, University of Oxford, UK): A 566bp exon-spanning region of human *LGR5* mRNA was amplified by PCR from human FAP polyp cDNA using specific primers and cloned into the pGEM®-3Z vector (Promega, UK), that has SP6 and T7 RNA polymerase primers flanking the multiple cloning sites, using the *Eco*RI

and *Hind*III restriction enzymes, then inserted into non-competent JM109 cells (Promega, UK). Once received, plasmids were then linearized with *Eco*RI and *LGR5* riboprobes synthesised with Sp6 RNA polymerase using Sulphur-35 (S^{35}) labelled uridine-triphosphate (UTP).

2.7.2 Tissue preparation

All PBS, ethanol and xylene solutions were pre-treated with 0.1% v/v diethylpyrocarbonate (DEPC) in order to inactivate any RNases. FFPE sections were cut at 4 μm, dried overnight at 40°C then de-waxed in fresh xylene for 8 minutes with two changes. Sections were rehydrated through 100%, 80%, 60% and 30% ethanol and then washed in PBS. Sections were permeabilised using Proteinase K (at a final concentration of 20 micrograms per millilitre in PBS) for 10 minutes at 37°C, rinsed in PBS containing 0.2% (w/v) glycine for 5 minutes and then PBS for 5 minutes. Slides were then post-fixed in 4% para-formaldehyde in PBS for 10 minutes, followed by rinsing in PBS (5 minutes x3). Tissue was acetylated with 500 millilitres of 0.1M triethanolamine + 1.25 millilitres of acetic anhydride added just before immersing the slides, then washed in PBS (5 minutes x3). Sections were dehydrated through graded alcohol containing 0.1% DEPC from 30% to 100% then air-dried.

2.7.3 Hybridisation

Hybridisation buffer was prepared as per Table 2.1. The final volume required was based on 10-25 microlitres per section, depending on the size of the section and the cover-slip to be used. Hybridisation buffer containing the *LGR5* probe was heated at 80°C for 1 minute, spun to reduce aerosols and pipetted onto the sections. Slides were then cover-slipped and placed into a humidified plastic slide-mailing box containing blotting paper saturated with 1x salts and 50% formamide. These were then put inside a lunch box humidified in the same way, sealed with PVC tape and incubated overnight at 55°C.

Proportion	Volume for 1ml	Ingredient
10%	100µl	10x salts mix in Denhardt's solution.
50%	500µl	Formamide
3%	30µl	rRNA
20%	200µl	Dextran Sulphate
1%	10µl	1 molar DTT
16%	160µl	Probe + DEPC treated H_20

Table 2.1 Hybridisation buffer mastermix
Example volumes are given for a 1 millilitre (ml) final volume; µl = microlitres.

2.7.4 Post hybridisation washes

Cover-slips were removed from each of the slides and placed in a clean rack, then immersed in 500 millilitres of 50% formamide buffer at 55°C on a rocking table for 3-4 hours; the formamide was replaced 3 times over the incubation period. Slides were then washed in 500 millilitres of TNE buffer at 37°C (5 minutes x9) and incubated with 100 micrograms per millilitre RNAse A in TNE buffer at 37°C for 1 hour in a plastic lunch box. A further 2 washes were then done in 2x SSC for 30 minutes at 65°C with agitation. Following this sections were taken through graded ethanols, increasing from 30% to absolute ethanol, all containing 0.3molar ammonium acetate to ensure that the labelled hybrids stay in place, then air-dried overnight.

2.7.5 Autoradiography

All steps were performed in a darkroom using a 902 filter and a 15-watt bulb. A 40 millilitre volume of Ilford K5 emulsion (15 millilitres stock emulsion + 25 millilitres 0.1% v/v DEPC treated H_20) was heated to 45°C in a water bath for 5 to 10 minutes. The solution was mixed using a glass rod then left for 10-20 minutes. Once the emulsion was judged to be of the correct consistency using a test slide, slides were

dipped into the emulsion so that the section was completely covered, placed on a cold metal surface and allowed to dry for about 2.5 hours until the surface of the emulsion had hardened. After drying, slides were placed in a wooden slide rack holder (or plastic) and silica gel added. The lid of the rack was then sealed with masking tape, inserted into a black plastic bag, sealed again with tape before refrigerating (+4°C) to expose the emulsion (7-14 days as appropriate).

2.7.6 Developing

Slides were immersed in pre-prepared Kodak D-19 developer at ~18°C for 4 minutes, agitated for one minute before washing in 1% acetic acid for 30 seconds followed by tap water for 30 seconds and then 30% sodium thio-sulphate fixer for two changes of 4 minutes each. Sides then underwent a further 10x 5 minute washes in tap water.

2.7.7 Counterstaining

Slides were immersed for 30-60 seconds in dilute stock giemsa solution (1: 100 with distilled water), washed under tap water and then air-dried before mounting under glass cover-slips in DPX. Sections were then examined under conventional or reflected light dark-field conditions (Olympus BH2 with epi-illumination) that allowed individual auto-radiographic silver grains to be seen as bright objects on a dark background.

2.8 Reverse-transcriptase polymerase chain reaction (RT-PCR)

In order to phenotype the clonal cell populations that had been identified by staining for CCO-activity, a technique involving CCO-histochemistry combined with RT-PCR for lineage specific mRNA was developed.

2.8.1 CCO histochemistry, RNA extraction and cDNA synthesis protocol

A protocol was devised to extract RNA from fresh frozen tissue, in combination with enzyme histochemistry for CCO-activity, by combining various elements from previously published RNA-extraction methods [165-167] and Zeiss online protocols (http://www.zeiss.de/labs):

1. All work surfaces and lab equipment was treated for RNase contamination with RNaseZAP (Ambion, Applied Biosystems, Warrington, UK).
2. P.A.L.M slides were treated for any RNase contamination by dipping in RNaseZap followed by 2 washes in DEPC-treated (0.1% v/v) H_20. Slides were then air-dried at 37°C for 2 hours and exposed to ultraviolet radiation (254 nanometre wavelength) for 45 minutes to improve section adhesion.
3. Serial sections of fresh frozen human colon were cut in a cryostat at -20°C onto the pre-treated P.A.L.M slides and allowed to air-dry at -20°C for 20 minutes.
4. In order to limit the time sections were exposed to the aqueous phase and, therefore, endogenous RNases, enzyme histochemistry for CCO-activity only was performed as previously described in section 2.4.1. Counter-staining for SDH activity was not done, with absence of CCO-staining used as a marker for CCO-deficiency. All staining solutions, PBS and ethanol solutions were pre-treated for RNases using 0.1% v/v DEPC.
5. Slides were dehydrated and fixed in graded ethanol: 75% for 1 minute; 95% for 1 minute; 100% 2x 1 minute. After air-drying for 1-2 minutes, single crypts were micro-dissected as previously described (see 2.2.2).
6. RNA-extraction was then performed using the Arcturus® PicoPure® RNA isolation kit (Applied Biosystems, Arcturus® products, Life Technologies Ltd, Paisley, UK). The manufacturers instructions were followed: An extraction buffer was pipetted onto the cap of the micro-dissection tube containing the micro-dissected tissue, and then incubated for 30 minutes at 42°C; the extract was then spun through a purification spin column to capture the RNA, followed by centrifugation with a series of wash buffers before the RNA was isolated using a low ionic strength elution buffer.

7. RNA-extraction was immediately followed by a reverse-transcriptase reaction in order to produce complimentary-DNA (cDNA) that could be stably stored at -20°C before PCR. Reverse transcription was performed using the SuperScript® III reverse transcriptase kit (Invitrogen, Paisley, UK) according to the manufacturers instructions. To summarise: mRNA was first specifically primed with oligo-dTs that bind to the 3' polyA-tail, following this the primed mRNA was incubated with a cDNA synthesis mix containing reverse transcriptase to produce cDNA template that could then be directly amplified using PCR.

2.8.2 PCR design and optimisation

As previously described, the human colonic crypt contains terminally differentiated cells of either the secretory or absorptive lineage. Therefore, in order to demonstrate the multi-potentiality of CCO-negative stem cell clones, primers specific for MUC-2, Chromogranin A, CD-10 and VILLIN mRNA were designed using the freely available *Primer-BLAST* (http://www.ncbi.nlm.nih.gov/tools/primer-blast) tool and sourced from Sigma-Aldrich, Poole, UK. All first-round primer pairs were exon-exon spanning to increase specificity for mRNA rather than DNA, and separated by at least one intron so that any DNA that was non-specifically amplified could be clearly seen on the agarose electrophoresis gel. Nested second-round primer pairs were then designed, as the initial amount of cDNA template was to be of low concentration. The primer pairs were optimised and nested-PCR protocols developed using control normal human colonic epithelial cDNA (see Appendix 9.1 and 9.2 for primer sequences and thermal cycler protocols). The cytoskeletal protein Beta-Actin was used as an internal positive control.

2.8.3 RNA extraction using Trizol® reagent to produce positive control cDNA

In order to produce positive-control mucosal cDNA to enable optimisation of the RT-PCR primers, a 0.5 centimetre x 0.5 centimetre piece of normal human colonic mucosa

was homogenised in 1 millilitre of Trizol® reagent (Sigma-Aldrich, UK). RNA was then isolated and cDNA synthesised using the following protocol:

1. 200 microlitres of chloroform was added and the tube shaken vigorously by hand for 15-20 seconds.
2. Samples were centrifuged at 13,000xg for 5 minutes to separate the organic and aqueous phases, and the colourless upper aqueous phase was carefully removed by pipette.
3. 500 microlitres of isopropanol was added and the sample incubated at room temperature for 15 minutes, followed by centrifugation at 13,000x g for 8 minutes to precipitate RNA.
4. The supernatant was removed by pipette to leave the RNA precipitate/ pellet, which was then washed once with 500 microlitres of 70% ethanol. The sample was mixed by vortexing, centrifuged at 13,000x g for 8 minutes, supernatant removed and then the RNA pellet air-dried for 3-4 minutes.
5. The RNA pellet was re-suspended using ddH_20 and reverse transcription performed using the SuperScript® III reverse transcriptase kit as detailed above.

2.9 Proteomics

It was hypothesised that matrix-assisted laser desorption ionisation mass spectrometry analysis (MALDI-MS) for lineage-specific proteins might enable gene expression analysis of CCO-negative clones within crypts. In order to assess the sensitivity of the technique for the small quantities of tissue involved, varying numbers of whole crypts were LCMD and analysed using the following protocol:

1. Serial sections of frozen, normal human epithelium were stained for CCO activity; Counter-staining for SDH activity was not done, with absence of staining for CCO used as a marker for CCO-deficiency.
2. LCMD of whole CCO-proficient and CCO-deficient crypts was performed. 14 microlitres of 80% ammonium bicarbonate/20% Acetonitrile was added as a lysis buffer and samples were then sonicated on ice.

3. The whole crypts were subjected to an overnight in solution tryptic digest (protein ratio 20:1) in 10 millimolar ammonium bicarbonate at 37°C. The supernantant was then dried down and washed prior to reconstitution in 5 microlitres of 50% (v/v) acetonitrile and 0.1% (v/v) trifluoroacetic acid.
4. 0.3 microlitres of the supernatant digest was spotted on a MALDI target plate and mixed with 0.3 microlitres matrix stock (10 milligrams per millilitre alpha-cyano-4-hydroxycinnamic acid + 1 microlitres 1 molar ammonium monophosphate to suppress matrix adducts in 0.1% trifluoroacetic acidin acetonitrile/water 50:50.)
5. The matrix solution was left to crystallize by evaporating the solvent and the sample analysed with a 4700 Applied Biosystems MALDI-TOF/TOF (matrix-assisted laser desorption ionisation tandem time-of-flight).

The mass spectra were initially advanced baseline corrected with a peak width of 32, flexibility of 0.5 and Degree of 0.1 using DataExplorer® (Version 4.6) software. A peptide "hit" was fragmented by high-energy collision-induced dissociation (CID). The resulting mass spectra of peptide masses were searched against the National Centre for Biotechnology Information (NCBI) non-redundant database (2010) using a Mascot search engine (Matrix Science Ltd) with a mass accuracy of 20 ppm and fragment tolerance of 100 ppm. One missed cleavage per peptide with the possible modifications of peptide N-terminal Gln to pyroGlu, oxidation of Met, and protein N terminus acetylated were considered in each search. Unidentifiable hits were analysed manually.

2.10 Methylation analysis

2.10.1 Laser capture micro-dissection and DNA extraction

Laser capture micro-dissection (LCMD) was performed as previously described (see Methods 2.2.2). Whole crypts from both normal colon and polyps were LCMD for methylation analysis; at least 3 serial *on-edge* sections were micro-dissected for each

crypt. For CCO-deficient clones within crypts, the blue and brown cell populations were separately micro-dissected from all serial *en-face* sections of the crypt into two laser-capture tubes.

DNA was extracted using 14 microlitres of proteinase K digestion buffer (Arcturus Bioscience, USA) for a minimum of 3 hours digestion. A negative control tube containing only digestion buffer and no laser capture material was also included. After digestion, tubes were briefly centrifuged and proteinase K deactivated by heating to 95˚C for 10 minutes. DNA lysate was stored at -20˚C.

2.10.2 Bisulphite treatment

For methylation analysis, extracted DNA was bisulphite treated using the EpiTect® Plus Bisulphite Kit (Qiagen, UK) according to the manufacturer's instructions: A sodium bisulphite mix is prepared and run on a thermal cycling programme overnight; the mix is then spun through an EpiTect® spin column with a buffer to promote binding of the converted DNA to the column membrane; bound DNA is then washed to removed residual sodium bisulphite, de-sulphonated, eluted from the spin column into a 1.5 millilitre micro-centrifuge tube and then stored at -20˚C until further use. An aliquot of the digestion buffer was also bisulphite treated as a negative control. Bisulphite treatment results in the efficient conversion of un-methylated cytosine to uracil, however it yields low DNA concentrations, thus in order to successfully PCR amplify the bisulphite treated DNA a two-round, nested PCR was used. CpG islands on the following genes were amplified: *MYOD1* (Myoblast determination protein 1, Chromosome 11p), *CSX* (Cardiac-specific homeobox, Chromosome 5q) and *BGN* (Biglycan, Chromosome Xq). These genes show limited or no expression in the colonic epithelium, age related methylation, and have been previously suggested for use as clonal markers in human colon crypts [18]. Specific primers flanking the target CpG islands for each gene were designed using the Primer 3 website (MIT, Cambridge, Massachusetts, USA), and reaction conditions (reagent concentrations and annealing

temperatures) were pre-optimised for each primer pair using cell-line DNA. Successful PCR amplification was confirmed by gel electrophoresis: only samples with an uncontaminated negative control were taken forward for cloning. For details of methylation specific primers, PCR programmes and post-bisulphite treatment CpG island sequences see Appendix 9.1, 9.2 and 9.3.

2.10.3 TA-cloning of PCR products

To achieve single strand resolution, PCR products were cloned using a TA-cloning kit, utilising the pGEM®-T vector (Promega, UK), according to the manufacturer's instructions. Taq DNA polymerase adds a terminal 3'-A overhang to the end of each PCR product, this enables it to be directly cloned into a linearized cloning vector that has single base 3'-T overhangs - *TA cloning*. The pGEM®-T is a linearised vector with a single 3'-T overhang at each end, and contains a *LacZ* gene and β-Lactamase coding region. PCR products are cloned in a ligation reaction containing PCR product, buffer, vector and T4 DNA ligase. The recombinant vectors are then transformed into competent JM109 cells (Promega, UK) and colonies grown overnight on agar growth medium containing ampicllin, X-galactose (X-gal) and Isopropyl β-D-1-thiogalactopyranoside (IPTG). Incorporation of the PCR product into the vector disrupts the *LacZ* operon so that successfully transformed clones are easily identified by blue/white colour discrimination. White clones were PCR amplified using M13 or T7/SP6 primer pairs that flanked the insertion point within the plasmid. The PCR product was then purified and sequenced as previously described (see Methods 2.3.2).

2.11 Statistical analysis of methylation patterns

Statistical analysis of methylation patterns was performed as formulated by Shibata and colleagues [18, 44, 168], with an additional measure – the *minimum distance* formulated by Dr Trevor Graham (Histopathology Lab, London Research Institute, Cancer Research

UK, 44 Lincoln's Inn Fields, London, UK). The pattern of methylated and un-methylated sites observed in a single strand (at a single locus) is described as a *tag*, with k=1,..,M CpG sites. For each crypt analysed from each patient, S_{jk}^{l} denoted the methylation status at locus l (*BGN*, *CSX*, or *MYOD1*), in the tag j, at site k. $S_{jk}^{l}=0$ denoted that the site was un-methylated, whereas $S_{jk}^{l}=1$ denoted the site was methylated. To describe the methylation present at each locus, in each crypt, three summary statistics were used:

a) *Proportion of methylated sites* (a proxy for mitotic age of the tissue).

This was calculated as:

- Percent methylated = $p_m = \frac{1}{NM}\sum_{j=1}^{N}\sum_{k=1}^{M} S_{jk}^{l}$

where N was the number of sequences analysed in that crypt at locus l, and M was the number of CpG sites at locus l.

b) *Intra-crypt* or *intra-clone distance* (a measure of epigenetic diversity within a crypt or clone). The distance between any two tags was defined simply as the number of sites where the methylation pattern differed (the Hamming distance). For example, the distance between the two sequences 1000110 and 0000111 is two, as they differ at the first and last CpG sites. The intra-crypt distance was the average distance between all tags from a crypt, calculated as:

Intra-crypt distance = $A = \frac{1}{N(N-1)}\sum_{i=1}^{N}\sum_{j=i+1}^{N}\sum_{k=1}^{M}\left|S_{ik}^{l} - S_{jk}^{l}\right|$

c) *Number of unique tags* in the crypt (a measure of epigenetic diversity within a crypt), denoted as U.

The similarity of methylation patterns between crypts was also assessed. The *inter-crypt distance*, the average distance between each tag in the first crypt compared to each tag in the second crypt was computed as:

- Inter-crypt distance = $I = \frac{1}{N_1 N_2} \sum_{i=1}^{N_1} \sum_{j=1}^{N_2} \sum_{k=1}^{M} \left| S_{ik}^{l} - C_{jk}^{l} \right|$

Where S_{jk}^{l} and C_{jk}^{l} represented the methylation tags from the first and second crypts respectively. N_1 and N_2 were the respective numbers of tags sampled from each crypt.

Second, the *minimum epigenetic distance* between crypts, which indicates the average number of methylation events required to transform each methylation tag in crypt 1 to those in crypt 2, was computed as:

- Minimum epigenetic distance = $E = \frac{1}{N_1 N_2} \sum_{i=1}^{N_1} \left[\min_{j=1}^{N_2} \left(\sum_{k=1}^{M} \left| S_{ik}^{l} - C_{jk}^{l} \right| \right) \right]$

The above statistical measures were designed into a Microsoft Excel spreadsheet by Dr Graham to enable analysis of methylation patterns both within and between crypts. In addition, a bespoke statistical program that allowed rapid analysis of the methylation pattern data was written using *Perl* programming language by Dr Graham. The respective statistical measures were then compared between two groups using the non-parametric Mann Whitney U-test and graphics produced using the freely available R Console platform (version 2.13) (http://www.r-project.org).

2.11.1 Mathematical Model

A mathematical model of colon crypt cell dynamics was used to examine methylation pattern divergence between daughter crypts following fission. The model was based on that of Yatabe and colleagues [18], but was modified in conjunction with Dr Trevor Graham to additionally describe crypt fission.

Each crypt was assumed to contain a constant *N* stem cells that divided at the same time as each other. At each round of division, each stem cell could produce either 0, 1 or 2 stem cell offspring, and a corresponding 2, 1 or 0 differentiated stem cell offspring, with probability p_0, p_1 or p_2, respectively (where $p_0+p_1+p_2=1$). Divisions were additionally constrained so that exactly *N* stem cells remained in the crypt following each round of division (for details see supplementary materials of [18]).

Methylation or de-methylation at each CpG site was assumed to occur independently and at a basal rate of $\mu=2\times10^{-5}$ per CpG site per cell per division. CpG islands were assumed to be initially un-methylated. Each cells contained two alleles of the *CSX* and *MYOD1* loci, and single allele of the X-linked *BGN* locus. Methylation patterns were inherited by daughter cells at division.

Crypt fission occurred after a predefined number of divisions ($T_f > 0$). During fission, the stem cells from the parent crypt were divided between the two daughter crypts. Parent stem cells were divided according to two segregation schemes. In the first scheme, crypt fission was assumed to be *symmetric*, such that the parent stem cells were divided equally between the two daughter crypts. Each of these *N/2* stem cells then underwent a symmetric division to form a crypt composed of the usual *N* stem cells. In the second scheme, a single stem cell from the parent crypt formed one of the daughter crypts, the other inherited the remaining *N-1* stem cells. *N-1* symmetric stem cell divisions restored the full compliment of stem cells in the first crypt, and a single symmetric stem cell division of one stem cell restored the full compliment of stem cells in the second crypt. After restoration of the full complement of *N* stem cells, crypt behaviour was as described above. Repeated simulation of the model, $>10^5$ repeats per parameter set, was used to investigate model behaviour.

2.12 Somatic mutation, LOH and MSI analysis of human polyps and adenomas

Fresh frozen human polyps were collected as previously described (see Methods 2.1). Serial sections were cut at 20 microns onto laser capture slides, with every 5th section cut at 6 microns onto normal glass slides for haematoxylin and eosin staining to enable accurate histopathological assessment by an expert GI pathologist (Professor Sir N.A. Wright, Histopathology Lab, London Research Institute, Cancer Research UK, London). Sections were then stained for CCO-histochemistry as detailed in 2.4.1. Tissue macro-dissection, LCMD, DNA extraction, PCR and sequencing were all done as previously detailed (see Methods 2.2 and 2.3). For details of all primer sequences, PCR reagent and thermal cycler protocols see Appendix 9.1 and 9.2.

2.12.1 Somatic *APC* mutation screening

DNA lysate from needle macro-dissected tissue and individual crypts was screened for somatic mutations in the tumour suppressor *APC* gene by PCR. Tumorigenic mutations in *APC* tend to cluster in the region between codons 1250-1450, termed the mutation cluster region (MCR) [169]. The MCR has previously been divided into 12 overlapping regions, each of around 100-200 base pairs, to facilitate PCR amplification from paraffin tissue and low DNA concentrations [107] (see Appendix 9.1 and 9.2 for primers and PCR protocols). As the quality of DNA in frozen tissue is better preserved than paraffin-embedded tissue, with no fragmentation, two sets of primers that amplify the whole MCR were designed using the freely available *Primer-BLAST* (http://www.ncbi.nlm.nih.gov/tools/primer-blast) tool and sourced from Sigma-Aldrich, Poole, UK (see Table 9.6). PCR protocols were successfully optimized for each primer set (see Table 9.6).

2.12.2 Somatic *KRAS* mutation screening

Oncogenic *KRAS* mutations tend to occur at codons 12 and 13, and occasionally codon 61 [137, 170]. Therefore, *KRAS* mutation screening of DNA lysate from whole tissue scrapes and/or individual crypts was performed using an established nested, two-round PCR specific for this region of the gene (see Table 9.9).

2.12.3 Somatic *BRAF* mutation screening

BRAF mutations are localised to codon 600. Therefore, mutation screening of DNA lysate from whole tissue scrapes and/or individual crypts was performed using an established PCR specific for this region of the gene (see Table 9.10).

2.12.4 Somatic *TP53* mutation screening

Exons 5-8 of the *TP53* tumour suppressor gene, where the vast majority of somatic mutations in TP53 occur, were amplified from whole tissue scrapes and/or individual crypts using an established nested, two-round PCR specific for these exons (see Table 9.11).

2.12.5 Microsatellite instability

Microsatellite instability (MSI) was assessed for DNA lysates from needle macro-dissected adenoma and polyp tissue using a multiplexed PCR for the *BAT-25* and *BAT-26* mononucleotide repeats (see Table 9.13). These repeats have been shown to be highly sensitive and specific for detecting constitutional defects in the mismatch repair (MMR) genes *hMLH1*, *hMSH2*, *hPMS1* and *hPMS2* [171]. The BAT25 marker is a poly(A) tract located within the intron of the c-KIT oncogene, and BAT26 is a poly(A)

tract located in the fifth intron of hMSH2. As these repeats are located within the introns of genes, they are highly conserved and thus sensitive reporters of DNA replication errors. Primers were tagged at the 5' end with either a FAM or HEX fluorescent marker. Successfully amplified PCR products were then analysed on an ABI 3100 Genetic analyser using Genotyper® 2.5 software (Applied Biosystems).

2.12.6 LOH analysis

Highly polymorphic microsatellites, located on chromosomes 5q (*APC*) and 17p (*TP53*) were amplified from DNA lysates using a multiplex PCR kit (Qiagen, UK) (see Table 9.12). Amplification of constitutional (normal) epithelium and adenomatous crypts was done in parallel to control for PCR variability. Marker accession numbers, primer and reaction details are all detailed in Appendix 9.1 and 9.2. Primers were tagged at the 5' end with either a FAM or HEX fluorescent marker. Successfully amplified PCR products were analysed on an ABI 3100 Genetic analyser using Genotyper® 2.5 software (Applied Biosystems). LOH was then considered present if the area under one allelic peak was more than twice that of the other, after normalising the peak areas relative to the constitutional DNA.

2.13 LINE-1 CpG site methylation analysis

Long interspersed nucleotide elements (LINE-1 elements) are highly repeated human retrotransposon sequences, and constitute about 17% of the human genome. Approximately one-third of DNA methylation occurs in repetitive elements and LINE-1 CpG site methylation, as measured by Pyrosequencing® (Qiagen, UK), is well correlated with global DNA methylation level [172] and therefore can be used as a proxy marker for global DNA methylation.

DNA lysate from needle macro-dissected tissue was first bisulphite treated using the Epitect® Plus Bisulphite Kit (Qiagen, UK) according to the manufacturer's instructions, as summarised previously (see Methods 2.10.2). Bisulphite treated DNA was then amplified using the PyroMark® CpG LINE-1 PCR kit (Qiagen, UK) according to the manufacturer's instructions. The primers are specific for a 146kb fragment containing four CpG sites, and the reverse primer is biotinylated to enable isolation of single stranded template DNA for sequencing using streptavidin beads. Pyrosequencing® (Qiagen, UK) analysis, based on sequencing-by-synthesis technology, to quantify the methylation level of the four CpG sites was then performed, according to the manufacturers instructions, using the PyroMark® Q96 MD sequencer: the biotinylated PCR product was first immobilised to streptavidin sepharose beads before purification and denaturation to a single stranded product using the PyroMark® Q96 vacuum workstation; the PCR product was then annealed to the sequencing primer before transferring to the PyroMark® Q96 MD and running the Pyrosequencing® reaction; the nucleotide dispensation order was: ATCAGTGTGTCAGTCAGTTAGTCTG. DNA from normal crypts, when available, and polyp tissue were analysed in parallel to control for any PCR variability. Technical assistance for Pyrosequencing® was kindly provided by Miss Athina Vidaki and Dr David Ballard of the Forensic Haematology Unit, Blizard Institute, Barts and The London School of Medicine and Dentistry, London, E3 2AT.

The sequencing data was then used to quantify the global methylation level at the four CpG sites. Figure 2.2 shows a trace from one of the samples analysed. Relative heights of the C and T peaks at each CpG site, as used in previous studies [135, 149, 172] were then used to give a quantification of the methylation level.

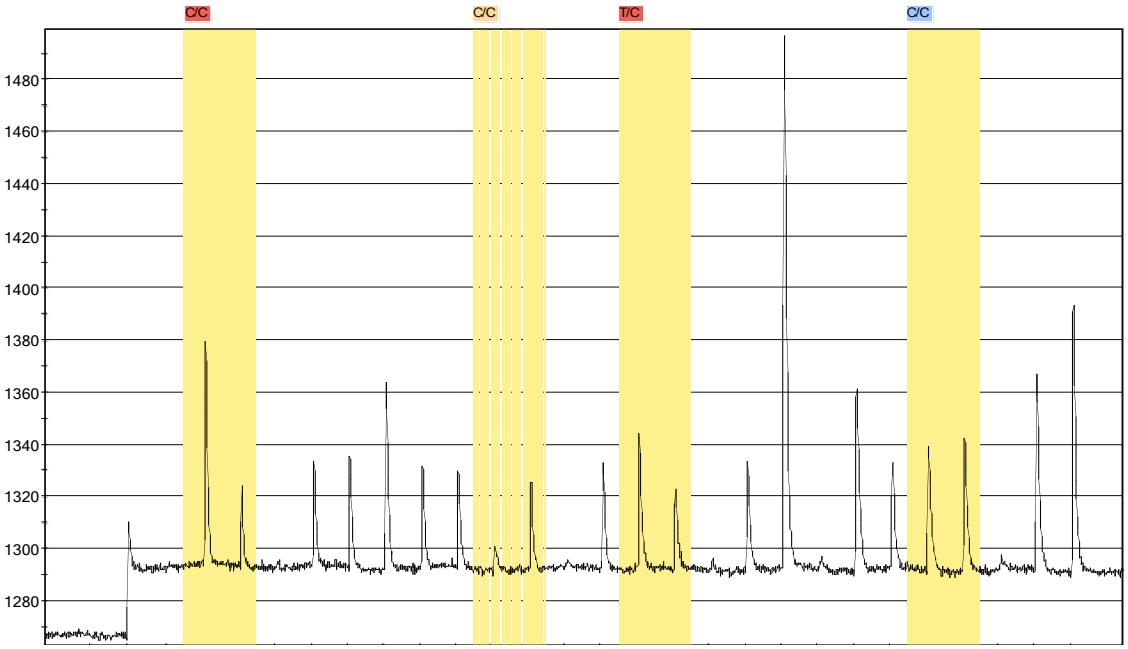

Figure 2.2 LINE-1 methylation as measure by pyrosequencing®
Example Pyrosequencing® trace. The four CpG sites within the LINE-1 sequence are highlighted in yellow. The relative ratio of the C and T peaks at the CpG sites is used to calculate the percentage methylation.

Chapter 3. Mapping stem cell clones in human intestinal crypts

3.1 Introduction and aims

Although it is now widely believed that there is a small number of related stem cells located at the base of the intestinal crypt that give rise to all the cell lineages of the crypt, most of the data to support this comes from mouse models due to the technical limitations of using human tissue. The aim of this project was to utilise naturally occurring, age-related somatic mtDNA mutations, that had previously been shown to accumulate in human colonic stem cells [76, 77] (see Introduction 1.1.6), as markers of clonality for whole human colon crypts. It was hypothesised that enzyme histochemistry and laser-capture sequencing combined with image analysis software could enable the mapping of long-lived, clonal cell populations within the crypts themselves and thus visualise putative stem clones.

As previously discussed (see Introduction 1.1.6. and Methods 2.4), after staining normal human colonic mucosa for activity of the mitochondrial encoded enzyme cytochrome *c* oxidase (CCO) three different types of crypts are observed: (i) wild-type, CCO-proficient brown crypts (ii) wholly CCO-deficient blue crypts and (iii) crypts containing populations of both brown and blue cells - mixed or partially-CCO-deficient crypts. Figure 3.1A-E demonstrates the different types of staining observed after CCO-histochemistry of fresh frozen human colonic mucosa. It is the mixed, partially-deficient crypts that provide a unique opportunity of identifying and mapping the progeny of a single stem cell lineage within individual human intestinal crypts.

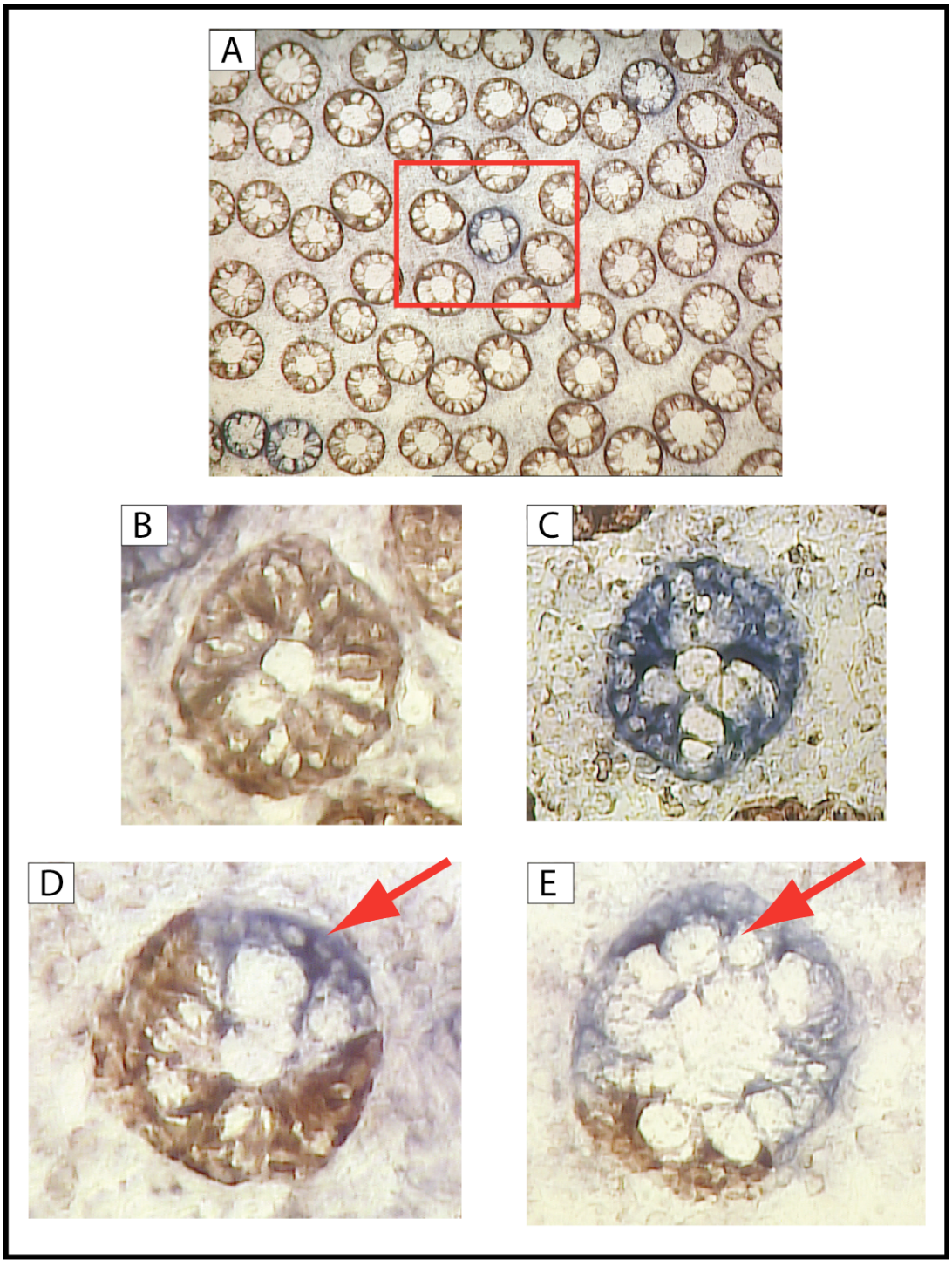

Figure 3.1 CCO-histochemistry of normal frozen human colonic epithelium
(A): 100x image of an en-face serial section after staining for CCO activity: most crypts have normal CCO activity and thus stain brown. However some crypts stain blue – see highlighted box - and are thus CCO-deficient. **(B):** 400x image of a normal, brown CCO-proficient crypt. **(C):** 400x image of a blue CCO-deficient crypt. **(D) & (E):** occasionally partially-deficient crypts are seen (400x), whereby a distinct population of blue cells is seen within the crypt (see red arrow). All sections were cut at 12 microns.

3.2 CCO-histochemistry identifies long-lived clonal cell populations within human colon crypts

In order to prove that these blue cell populations within crypts are indeed clonal and derived from a single stem cell lineage, individual cells from within the blue and brown cell populations of the crypt were laser capture micro-dissected and the mitochondrial genome amplified in order to detect clonal mutations in the blue cells (see Methods 2.4). Figure 3.2A demonstrates a colonic crypt that, after CCO-histochemistry staining of serial en-face frozen sections of human colon, contained a distinct population of blue cells that extended from the base of the crypt to the luminal surface. Sequencing the mtDNA of cells, from both within the partially-CCO-deficient crypt and neighbouring crypts, confirmed that all the blue cells within the crypt, identified by CCO-histochemistry, were a distinct clonal population (Figure 3.2B-D) and thus the progeny of a single stem cell lineage.

More complex mtDNA mutation patterns were observed within crypts. As previously discussed, the mutations in the mitochondrial DNA are acquired over a number of years by a stem cell via genetic drift [76], and this stem cell can then expand stochastically over time within the niche by niche succession. Once a high proportion of the mitochondria within the stem cell have the mutation, then this is enough to cause a deficiency in oxidative phosphorylation [81]. In Figure 3.3 a blue cell population is seen occupying a large proportion of a colon crypt. After sequencing, the blue cells were found to have a clonal mutation in the mitochondrial encoded NADH dehydrogenase 2 (MT-ND2) gene (C>A at position 4888) at a high heteroplasmic level - note a small peak of the wild-type base was also seen in the sequencing trace from the blue cells (Figure 3.3A-B). Upon sequencing the brown cells from within the crypt the mutation was also picked up, but at a lower, heteroplasmic level, whereas surrounding brown crypts where homoplasmic for the wild-type mitochondrial sequence (Figure 3.3A-B).

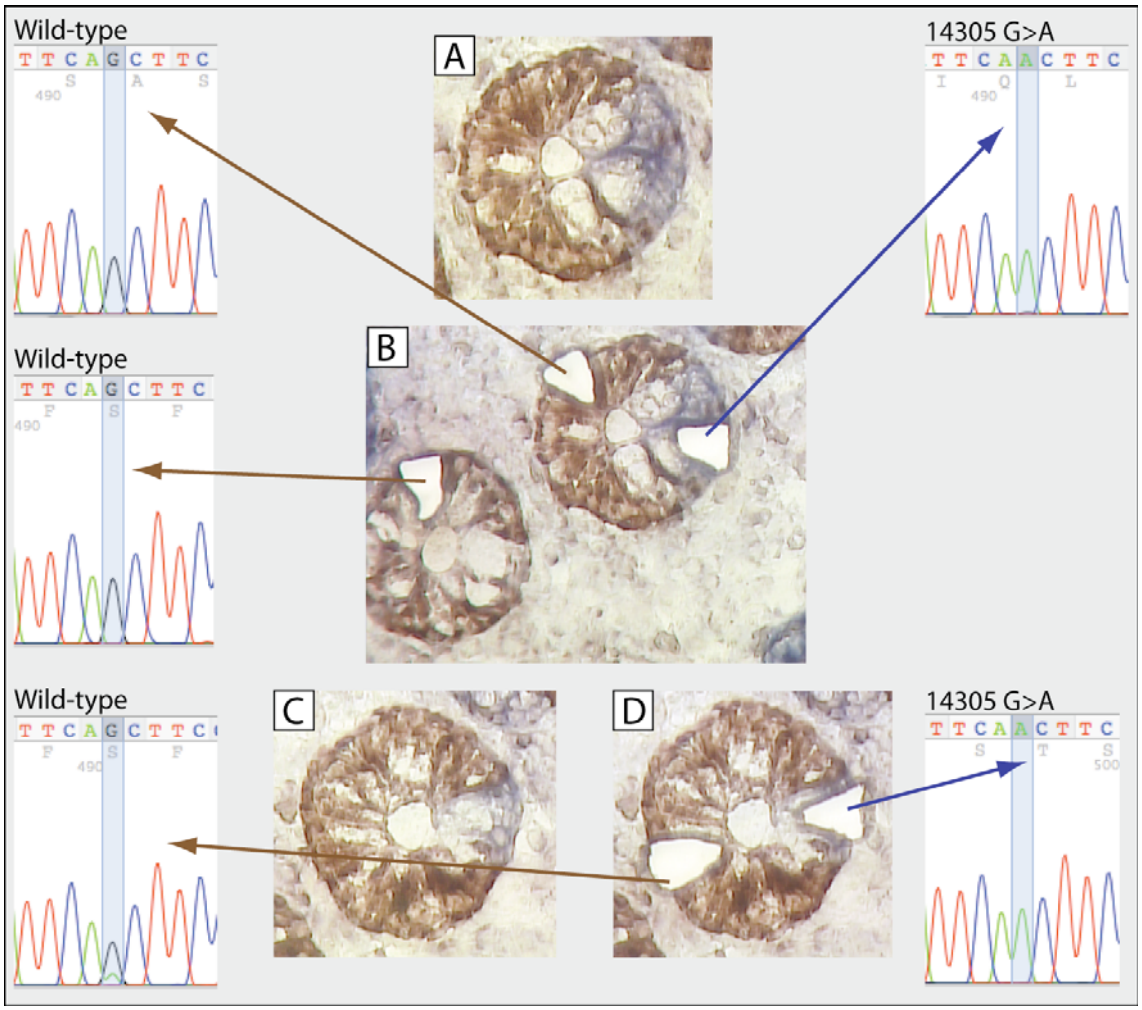

Figure 3.2 CCO-histochemistry identifies a clone from a single stem cell lineage within a human colonic crypt
CCO-staining of serial en-face sections allows the identification of stem cell clones within crypts. **(A):** serial section of a crypt containing a distinct population of CCO-deficient, blue cells. **(B):** LCMD of blue and brown cells within the crypt, and of brown cells from adjacent crypts demonstrates that the blue cell population contains a clonal mutation in their mtDNA, in this case a non-coding G>A at position 14305 in the MT-ND6 gene, whereas brown cells are wild-type. **(C)** & **(D):** LCMD of blue and brown cells from sections at different levels within the crypt confirm that all blue cells harbour the same clonal mutation. All images are at 400x magnification.

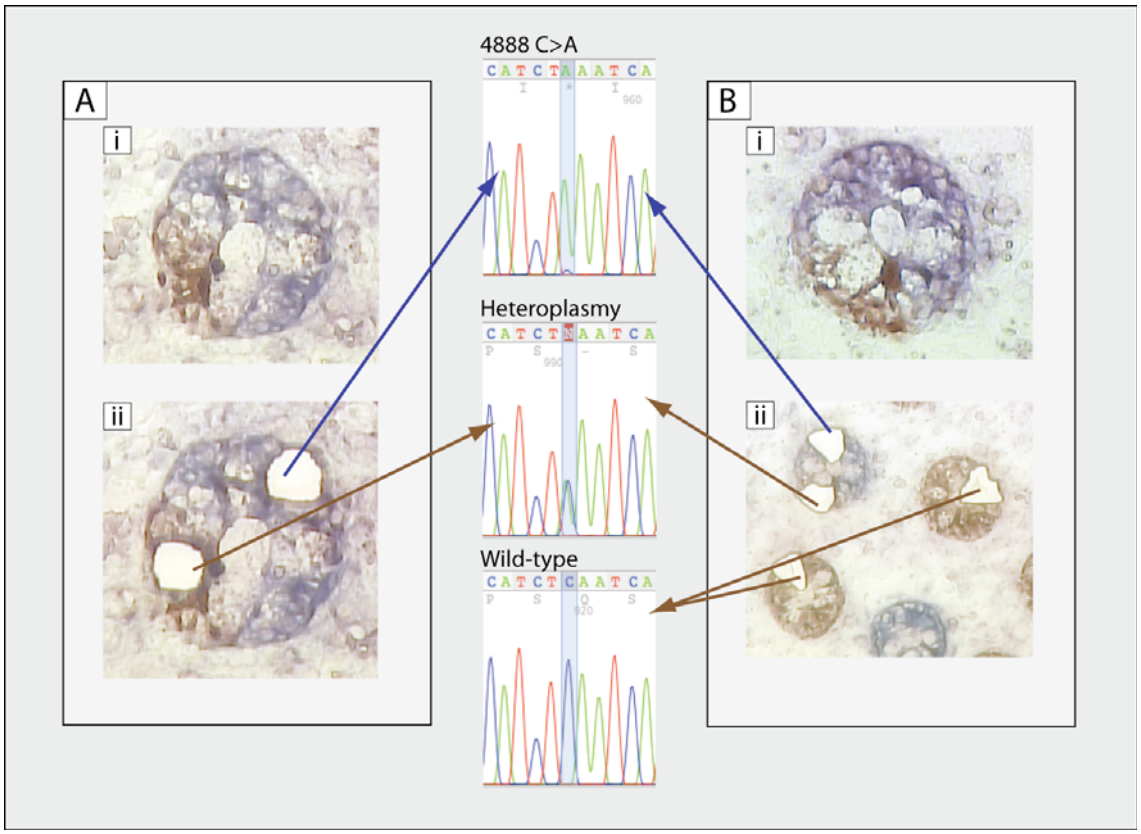

Figure 3.3 Crypt niche dynamics
(A): 400x images of an en-face serial section through a partially mutated crypt, (i) pre and (ii) post-LCMD. The CCO-deficient blue cells all contain a 4888 C>A mutation in their mtDNA that is also seen at heteroplasmic levels in the brown cells. **(B):** (i) 400x pre-LCMD image of an en-face serial section from a different level of the crypt; (ii) 200x post-LCMD image of same section - sequencing confirmed the clonal mtDNA mutation in blue cells of the crypt with brown cells again demonstrating heteroplasmy, whereas adjacent crypts were wild-type (ii).

For the crypt in Figure 3.3 the mtDNA mutation was located in the *MT-ND2* gene that encodes for the nicotinamide adenine di-nucleotide (NADH)-ubiquinone oxidoreductase chain 2 enzyme (complex 1), this gene is immediately upstream of the genes encoding for CCO and the point mutation resulted in a premature stop codon (TCA⇒TAA). As transcription of the circular mitochondrial genome occurs uninterrupted once it is initiated at the promoter region [173], it is clear how this mutation would cause a subsequent deficiency in the cytochrome *c* oxidase enzyme. However, as found by previous studies [76], it was not always possible to identify a mitochondrial mutation in the blue cell population of partially-deficient crypts. Figure 3.4 demonstrates such a crypt: there was a clearly defined CCO-deficient blue cell population (Figure 3.4A), but after LCMD and sequencing of the entire mitochondrial

genome of both blue and brown cells from within the crypt no clonal mutation was identified. However, upon checking the polymorphisms detected in the partially-mutated crypt by sequencing the mtDNA of cells from adjacent blue and brown crypts, one of the 'polymorphisms' detected in the partially-mutated crypt was found, in-fact, to be a clonal mutation for the entire crypt as no surrounding crypts harboured the same polymorphism (Figure 3.4B-D).

The mutation in this case, a T>C at position 152, is in the non-coding control region of the mitochondrial genome, which contains the promoter binding regions for the initiation of both heavy and light-strand mitochondrial DNA transcription [173]. It is thus possible that a secondary clonal epigenetic change that resulted in impaired transcription factor binding in the CCO-deficient blue cell population within the crypt was the cause of the detectable CCO-deficiency in this case. There are a number of technical reasons why clonal mutations in blue, CCO-deficient cells within the crypt might be missed: the mitochondrial genome is 16.6kb and, although the primers are overlapping, after sequencing there are occasionally small (<10bp) areas at the start or end of the amplicon, around the primer binding sites, that cannot be read; very small areas of tissue are being LCMD, perhaps 2-3 cells, and CCO staining intensity is variable so it is easy to get contamination from brown, wild-type cells; the PCR reaction may preferentially amplify the wild-type sequence in those clones that are heterozygous for the clonal mtDNA mutation. Importantly, never were blue cells from within clones found not to share a mtDNA mutation when one was identified, and no previous authors using this technique have reported this either [76-78, 174].

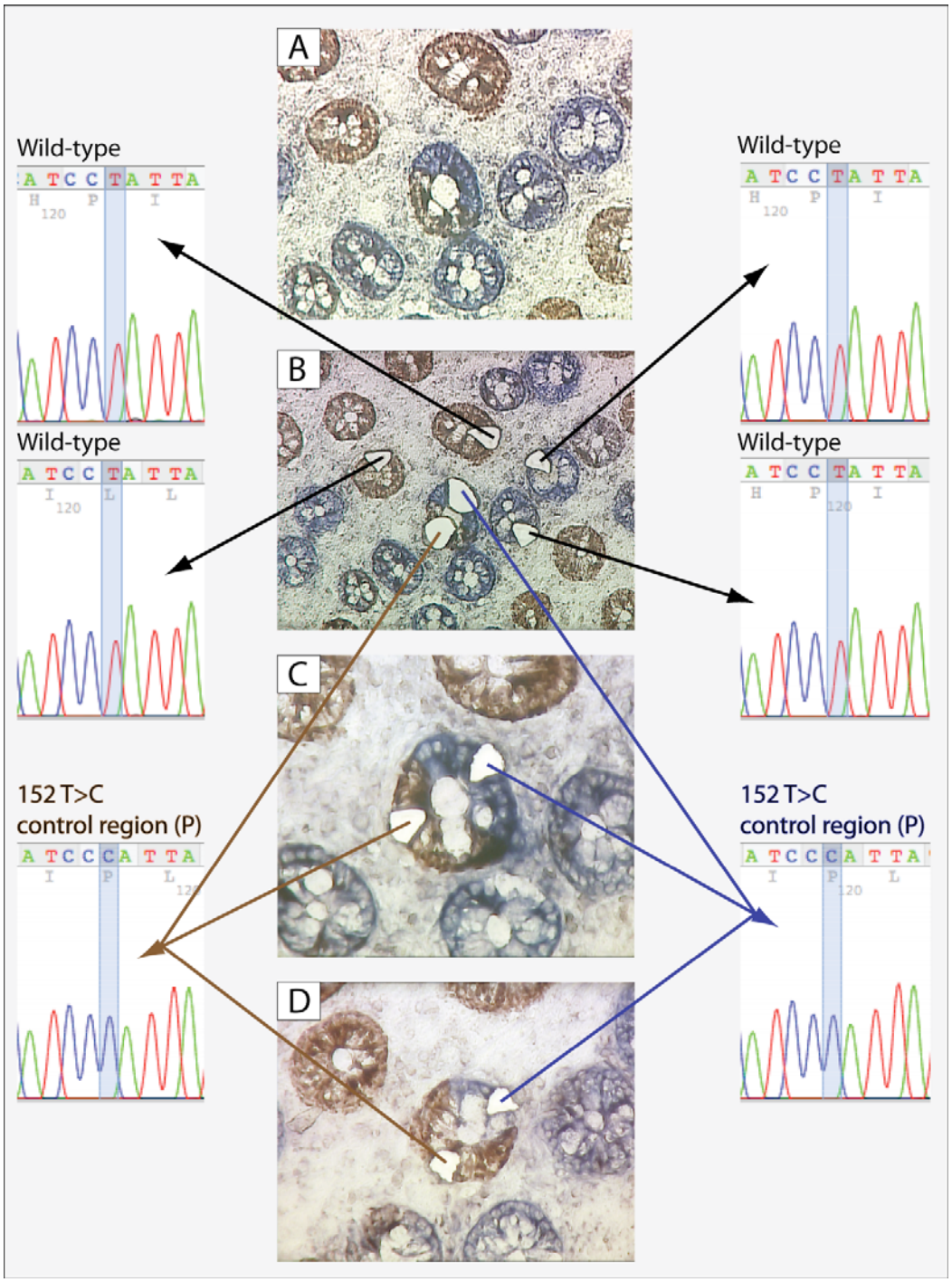

Figure 3.4 Partially CCO-deficient crypt with no clonal mtDNA mutation in the CCO-deficient cells

(A) & (B): pre (A) and post (B) LCMD of a partially CCO-deficient crypt and neighbouring blue and brown crypts demonstrates a clonal polymorphism (152 T>C) in the control region of the mtDNA of both CCO-proficient and CCO-deficient cells in the partial crypt; neighbouring blue and brown crypts are wild-type for this polymorphism (x200). **(C) & (D):** LCMD from sections at two further levels of the partial crypt confirms the presence of the clonal polymorphism in all blue and brown cells of the crypt (x400).

3.3 Mapping of stem cell clones

In order to be able to visualise the path that these stem clones were taking within the crypts, a novel method of mapping the origin and spread of the mutated blue cell populations within the intestinal crypt was devised by combining the two-colour enzyme histochemical staining described with image analysis and computer reconstruction (see Methods 2.5), using bespoke software written by Dr Paul Tadrous (Histopathology Dept, Northwick Park Hospital, Harrow, London) (www.Bialith.com).

Figure 3.5 illustrates the generation of a 'crypt map' from multiple serial transverse images through the crypt, the resulting crypt maps are 2D images whereby the 3D tubular crypt has been cut open and laid flat to fully illustrate the position and path of the clone within the crypt. Serial en-face sections were cut at 8, 10 or 12 microns, depending on the quality of the frozen tissue. Only crypts that could be traced from the base to the surface were used for mapping. During tissue processing occasional sections were of poor quality and these were not subsequently used for constructing the maps. To map a single crypt from base to top usually required the processing of around 30-40 serial sections, depending on the size of the crypt. Once the serial frozen sections had been stained for CCO-histochemistry and a partially-deficient crypt identified, an individual map could be produced within 7 or 8 hours. The image analysis software produces three images: (i) *unsaturated*: the basic crypt map where none of the staining hues have been enhanced; (ii) *saturated*: here the natural staining hues have been enhanced; (iii) *blue/black*; in this image all blue cells remain blue but all non-blue staining cells appear black, this enables the origin and spread of the blue cells within the crypt to be clearly visualised.

In total, 32 colon crypt maps and 4 small bowel maps were generated from 11 anonymised patients who had all undergone surgery for either colorectal polyps or cancer; fresh frozen morphologically normal tissue was collected from outside the resection margins. The mapping software proved very consistent, as there was no variation in the appearance of the blue, CCO-deficient clones if the program was run

choosing different points along the crypt circumference at which to cut the crypt to obtain the 2-D image.

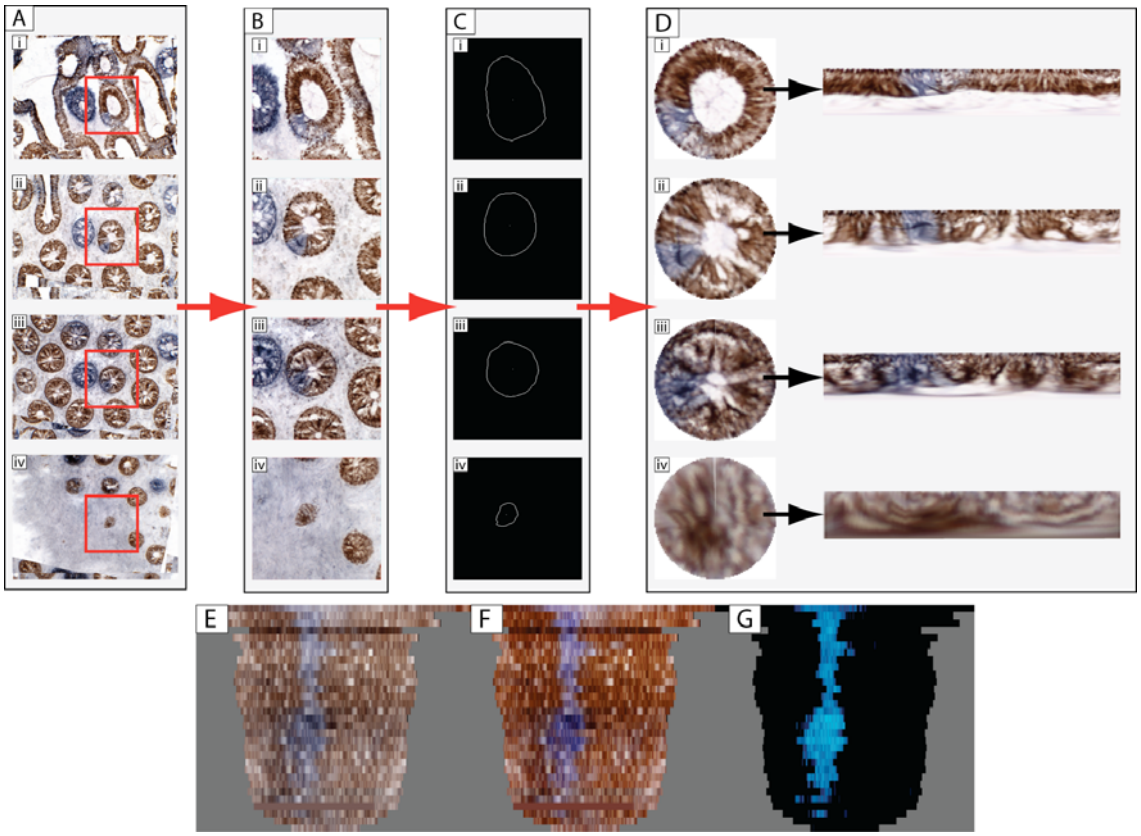

Figure 3.5 Generation of crypt map from a partially CCO-deficient crypt
Images **(i)-(iv)** are of the top, upper, lower and base of the crypt respectively. **(A):** original images of serial sections taken at 20x magnification on a light microscope that have been aligned relative to each other at a sub-pixel level using BiaQIm software (www.bialith.com); a partially deficient crypt to be mapped is highlighted by the red box. The target crypt is then cropped out **(B)**, and for each section the outside of the crypt is traced and the centre marked to produce mask images **(C)** for the mapping software. Bespoke software then builds the 2-D crypt maps using the serial cropped images and their respective masked profiles: **(D)** circular images are produced which are then cut at a chosen angle and warped into a rectangular profile. Each rectangular profile is then averaged to one scan line and stacked on top of each other to produce the crypt maps: **(E)** original 'unsaturated' crypt map; **(F)** 'saturated' map where the natural staining hues have been enhanced; **(G)** final blue/black crypt map that enables the path taken by the blue cells to be easily visualised.

3.3.1 Most CCO-deficient clones arise from the accepted area of the human colonic stem cell niche

Mapping of stem cell clones from human colonic tissue revealed a variety of different patterns of spread of stem cell progeny within colonic crypts. For the majority of clones it can clearly be seen that the first mutated cells appear towards the base of the crypt: Figure 3.6A-B show mutated clones arising just above the crypt base, expanding upwards in a linear fashion through the crypt to reach the top. In Figure 3.6C-E clones are shown that arise from right at the bottom of the crypt, in cell position 1. The crypt maps shown in Figure 3.6 are representative of most of the colonic stem cell clones observed in this study, with the clonal populations arising from the area of putative colonic niche, however there is lateral movement of cells as they progress up the crypt which is in contrast to what is observed in mouse lineage tracing studies [19].

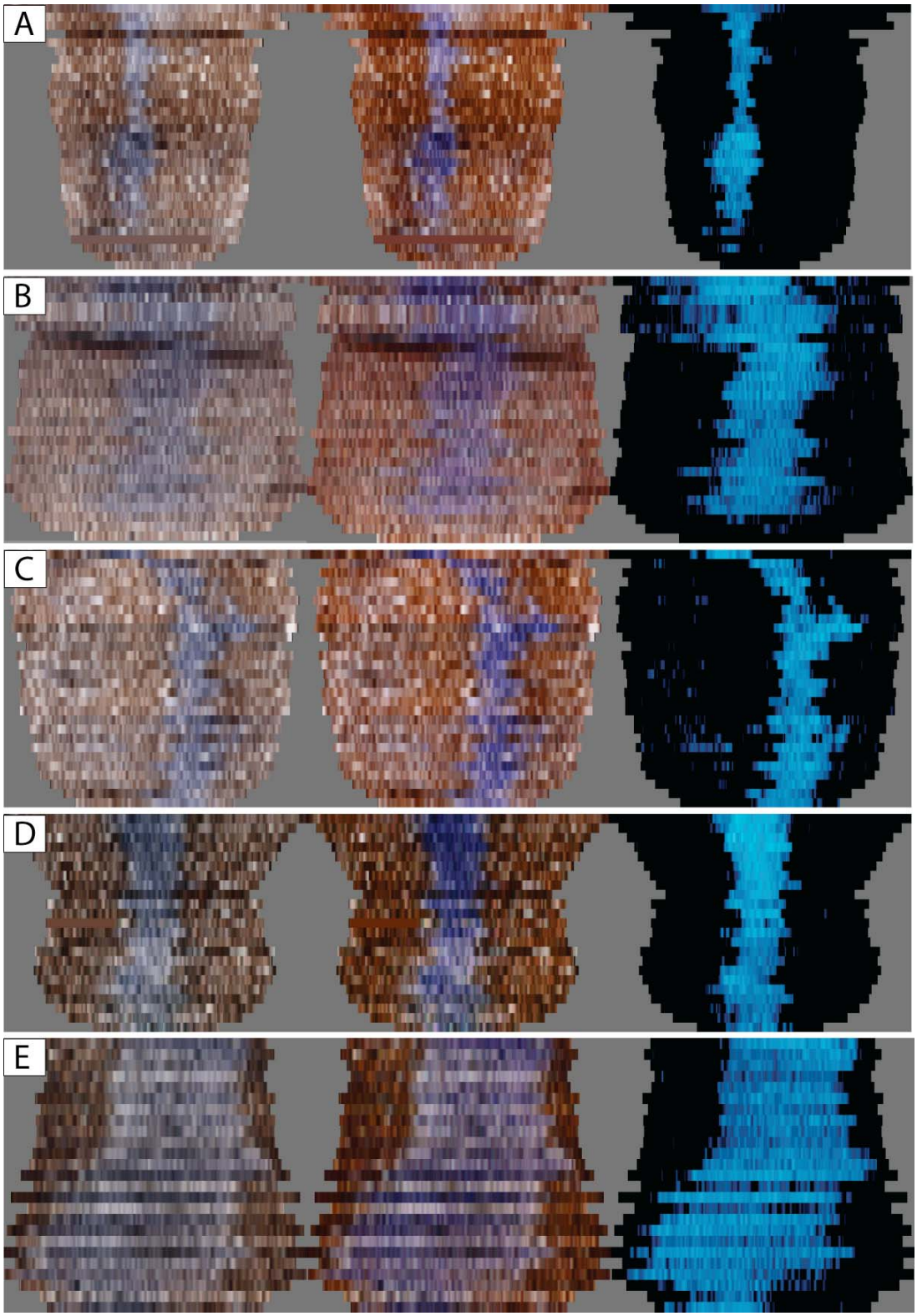

Figure 3.6 Crypt maps of human colonic stem cell clones
For each crypt the unsaturated, saturated and blue/black colour-processed maps are shown from left to right. **(A) & (B):** mutated clones are seen arising just above the crypt base, at cell position 2-3. **(C-E)**: mutated clones arising from the base of the crypt, this is consistent with the stem cell niche being at cell position 1-3 in the human colon.

3.3.2 Complex, atypical clones are seen within human colonic crypts

The pattern of clonal expansion was not always linear, as has been previously suggested [51], but clones were observed expanding laterally as they migrated up the crypt, with significant mixing of cells seen (Figure 3.7A-B). Occasionally clones arising from positions outside the accepted location of the stem cell niche were found: In Figure 3.7C-D, two clones are illustrated that appear to arise from approximately one-third of the way up the crypt, well above cell positions 1-3.

A small number of maps have shown what appear to be more than one mutated clone arising within the same crypt (Figure 3.8A-B), the blue cells could all be clonally related and we are simply seeing varying patterns of cell migration, or they could represent separate, independent stem cell clones providing evidence that there are at least 3 stem cells in the human colonic niche. Rarely clones that did not appear to reach the luminal surface of the crypt were seen (Figure 3.8B-C), and this could indicate the existence of more quiescent stem cells within the niche or variable stem cell cycling.

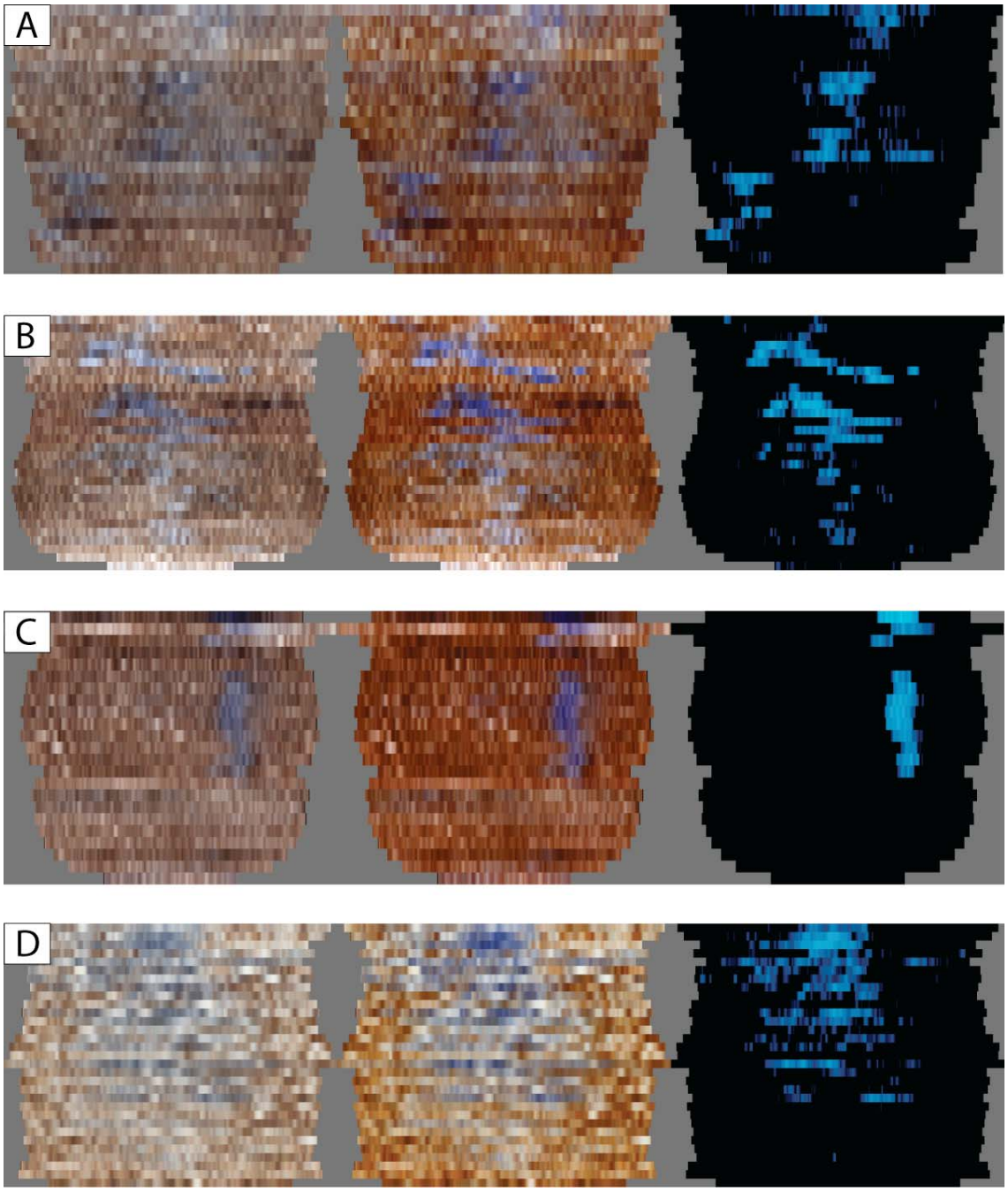

Figure 3.7 Colonic clones demonstrate different patterns of expansion within the crypt
Cell migration is not always linear: in **(A)** a clone is seen arising at the crypt base and spreading in a diagonal fashion toward the top. In **(B)** it can be seen that there is lateral mixing of cells during migration to the luminal surface. **(C) & (D):** Clones were observed arising from almost half way up the crypt, outside the accepted area of the colonic stem cell niche.

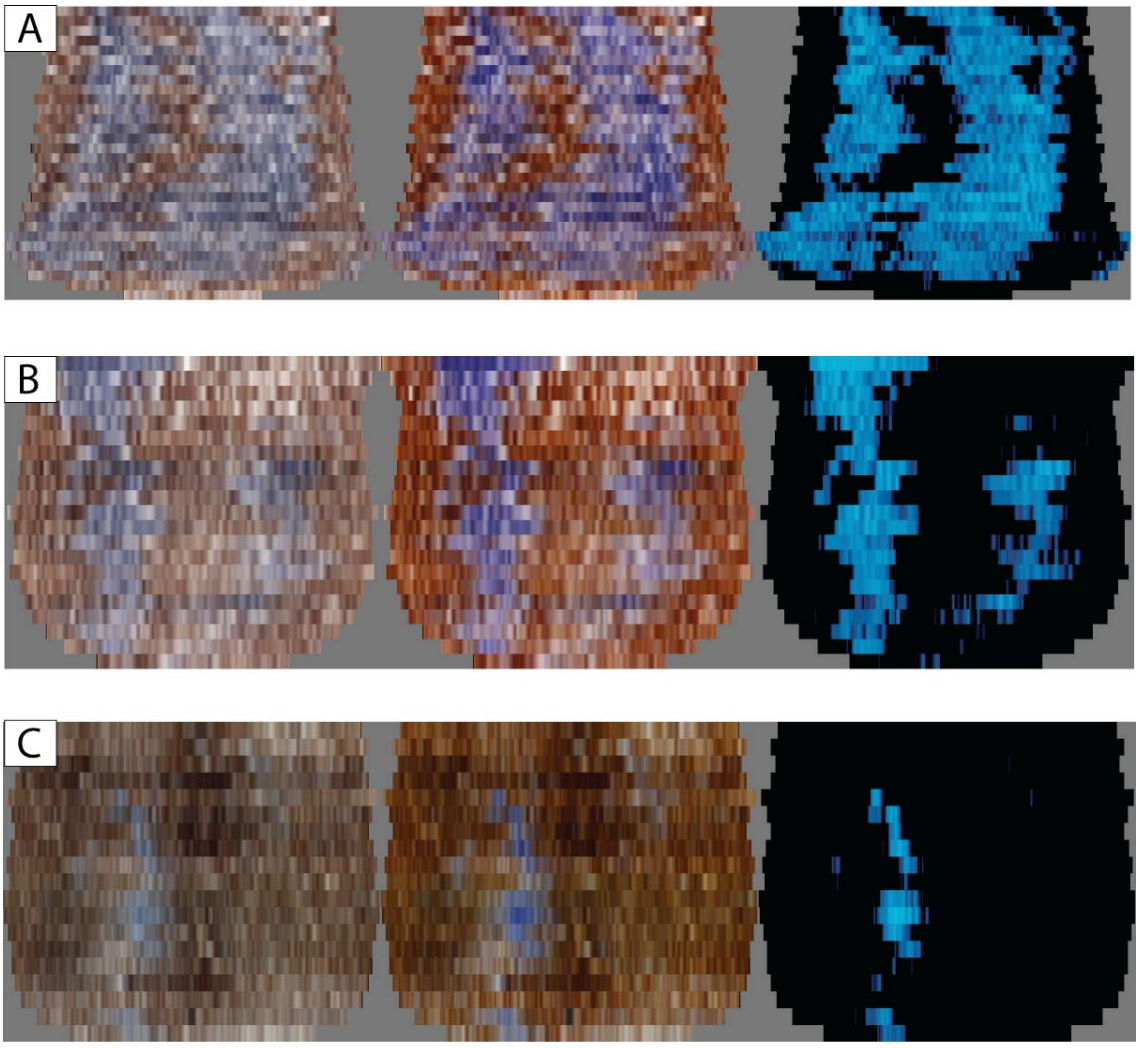

Figure 3.8 Multiple clones within single crypts and apparent relatively static cell populations
In a minority of crypts, what appears to be more that one mutated clone is observed arising from the stem cell niche **(A & B)**. Occasionally clones are observed that do not reach the top of the crypt (**B & C**), suggesting that there may be clonal cell populations within the crypt that are relatively quiescent or have variable cell cycling dynamics.

3.3.3 Small bowel crypt maps

The frequency of wholly CCO-deficient crypts observed in the small bowel is significantly less than the colon [79], and thus finding partially CCO-deficient crypts is very rare. Therefore, fewer crypts from the small bowel were able to be mapped. The small bowel crypt maps generated and shown in Figure 3.9 appeared to suggest that the location of the stem cell niche in the human small intestine may be indeed above the basal Paneth cells, around cell position 4-6, as has been suggested by previous mouse studies [20, 26, 175].

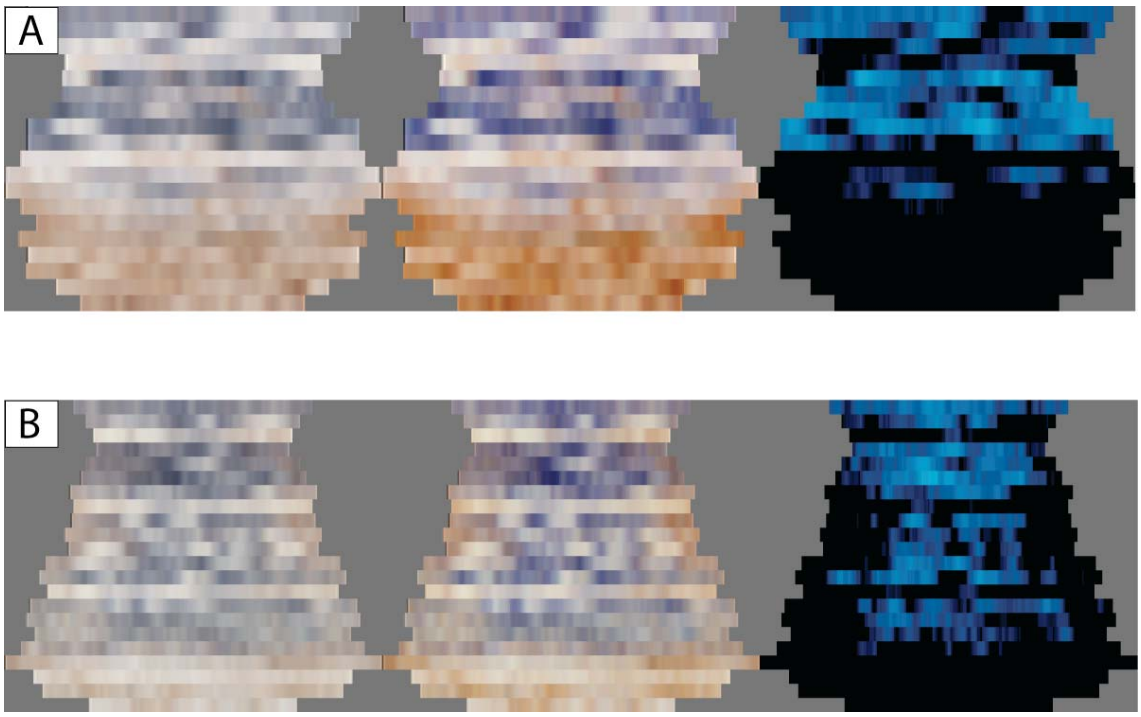

Figure 3.9 Human small bowel crypt maps
(A) & (B): Mapping of clones within human small intestinal crypts suggests that the stem cell niche is located in the bottom third of the crypt, above the basal Paneth cells.

3.4 Discussion

Taylor et al [76] first observed that, in elderly patients, a minority of human colon crypts appear deficient in the mitochondrially-encoded enzyme cytochrome *c* oxidase (CCO), and upon sequencing the mtDNA of cells from these CCO-deficient crypts they all harboured identical, clonal somatic point mutations in their mtDNA; they also noted the existence of CCO-deficient populations within crypts. The authors concluded that, since these mutations offer no selective advantage, only a stem cell lineage would have sufficient life-span to acquire these mutations over many years by genetic drift. This mutated stem cell must then stochastically expand within the niche, via niche succession, with subsequent monoclonal conversion resulting in all the progeny of that stem cell lineage populating the crypt causing it to appear CCO-deficient. These CCO-deficient crypts are not seen in patients less than 40 years old [77], so it is clear that a cell lineage would have to have a lifespan of at least 40 years to enable the accumulation of these mutations.

In this chapter, it has been shown that, by staining for the activity of CCO and sequencing the mtDNA of CCO-deficient and CCO-proficient cells within human colon crypts, long-lived clonal cell populations that represent a single stem cell lineage can be identified within crypts themselves. Most mtDNA mutations detected were homoplasmic i.e. all the mitochondria within the cell harboured the mutation. However, the observation of differing heteroplasmic levels of mtDNA mutations within CCO-deficient and CCO-proficient cells of the same crypt highlights the dynamics of the stem cell niche, whereby all stem cell lineages are related but some more recently than others. In Figure 3.3, a clonal mtDNA mutation was detectable in both the blue and brown stained cell populations, therefore, it appears that after a number of rounds of niche succession cycles all stem cell lineages within the niche had acquired this mutation to some degree. However, in the stem cell lineage supporting the CCO-deficient clone, the mutation has drifted to dominance and is at a high enough heteroplasmic level to cause a phenotypic CCO-deficiency. In the remaining stem cells of the niche, the mutation has remained at low heteroplasmic levels, thus their progeny are CCO-proficient. These results support the conclusions by Yatabe et al [18] that the

human colon stem cell niche contains multiple, related stem cells that undergo successive niche succession cycles, whereby a single stem and its direct progeny will inevitably come to take over the niche with extinction of all other stem cell lineages. The exact number of stem cells within the niche is not known, but technique has clearly demonstrated that the human intestinal crypts contain at least two stem cell lineages, and by successfully applying mtDNA sequencing for crypts containing two distinct blue cell populations, it may be possible to prove the existence of at least three stem cell lineages within the human colonic niche.

By applying image analysis software, it was possible to develop a method for reconstructing whole crypts containing CCO-deficient putative stem cell clones, in order that the origins and path taken by the clones could be fully visualised. The crypt maps in Figure 3.6 are representative of the vast majority of clones seen, indicating that, in the human colon, cells capable of forming sustained clones are present in cell positions 1-3 in the crypt, which corresponds to the current accepted site of the stem cell niche in the human colon. Clones were occasionally observed arising from approximately one-third of the way up the crypt, well above cell positions 1-3; these could represent mutated stem cell clones that are being flushed from the niche by clonal succession, or equally it could represent a mutation occurring in a long-lived committed progenitor cell. Indeed, by marking random intestinal epithelial cells by somatic mutation of the *Dlb-1* locus in mice, Bjerknes and Cheng [17] were able to demonstrate short-lived (days) progenitors yielding one or two cell types and long-lived (months) mucous cell progenitors and columnar (enterocyte) cell progenitors, as well as multipotential stem cells capable of giving rise to all epithelial cell types. Consequently, it is quite possible that the clones in Figure 3.7C-D represent a mutation in a committed mucous or columnar cell progenitor located in the mid-crypt, a proposal that could be tested by lineage analysis of the cells within these rare, atypical clones.

The progression of the progeny of stem cells from the base of the crypt to the luminal surface is generally thought to occur in a simple, straight, linear fashion, as this has been observed in mice [19, 20, 51]. However, crypt maps of CCO-deficient clones in human intestinal crypts have shown that clones arising from the base of a crypt can take a

variety of paths to the luminal surface, with some seen to take a diagonal route, and there can be significant lateral movement of cells as they move up the crypt, resulting in mixing of progeny from different stem cell lineages. This may be a consequence of greater dynamic physical forces upon cells within the human crypt, due to the increased numbers of cells compared to murine crypts.

Mapping of putative stem cell clones in the human small intestine suggests that the stem cell niche may indeed be around cell position 4, above the basal Paneth cells as has been identified in many murine studies [20, 26, 175]. The idea that the structure and biology of the niche may differ as one progresses from the small to the large bowel would be consistent with observed clinical phenomenon. Adenocarcinoma is very rare in the small intestine, yet the third commonest cause of cancer worldwide in the colon, with left sided colorectal cancer more common than proximal adenocarcinoma. Thus, it may be that this reflects the differing stem cell biology: clonal stabilisation times in the murine intestine are 4 weeks for the colon and 21 weeks for the small intestine respectively [10], and mutated CCO-deficient clones and crypts are much more frequently detected in the human colon versus small bowel [79] which may be due to their being significantly greater numbers of stem cells within the small bowel niche. A tumorigenic stem cell within a colonic stem cell niche with fewer stem cells would have less competition, and thus it would potentially be easier for it to take over the niche and become fixed within the crypt.

The mtDNA mutations observed in CCO-deficient clones were not always the obvious cause of CCO-deficiency, often being outside of the genes encoding for CCO and occasionally in the non-coding control region. However transcription of the mtDNA is not well characterised and is very different to nuclear DNA. The mitochondrial genome is circular and transcription, once initiated by the binding of transcription factors within the control region, occurs uninterrupted [173], thus mutations anywhere within the genome can have downstream effects. The epigenetic regulation of mtDNA transcription is also not well understood, the clonal mutation seen in the crypt in Figure 3.4, a T>C within the control region at position 152, may well have predisposed to a second, clonal epigenetic change that caused the CCO-deficiency within the blue cell population of the

crypt. It is also possible that, when no clonal mtDNA mutations were detected, there may have been a mutation in the nuclear genome involving other proteins that regulate mtDNA replication and transcription, or form part of complex IV in the respiratory chain, as suggested by Taylor et al [76].

.

Chapter 4. Delineating human colonic stem cell clones

4.1 Introduction and aims

The results from the previous chapter established a technique for identifying and mapping a single stem cell lineage and their direct progeny within crypts. The 'gold-standard' for identifying stem cell populations is demonstrating their multi-lineage potential in addition to self-renewal [176], and this has been established in recent mouse lineage tracing studies of the intestine [19, 20]. These elegant techniques employed in genetically engineering mice obviously cannot be used in humans, therefore the aim of this project was to phenotype CCO-deficient clones within human colon crypts in order to demonstrate multi-lineage potential.

4.2 Proteomic studies

It was hypothesised that mass spectrometry analysis for lineage-specific proteins might enable gene expression analysis of CCO-deficient clones within crypts. A protocol was developed, in collaboration with the Proteomics laboratory at The London Research Institute, Cancer Research UK, using matrix-assisted laser desorption ionisation mass spectrometry analysis (MALDI-MS) (see Methods 2.9). In order to first assess whether the technique was sensitive enough to enable analysis of protein expression from crypts stained for CCO activity, 10 and 5 whole crypts were micro-dissected for proteomic analysis. Only CCO-deficient crypts were micro-dissected as the diaminobenzadine (DAB) stain in CCO-proficient cells interferes with MALDI-MS analysis (Figure 4.1). No SDH counter-stain was performed in order to prevent interference with the MALDI-MS analysis, and CCO-deficient crypts were therefore identified by the absence of DAB (brown) staining (Figure 4.1A-B).

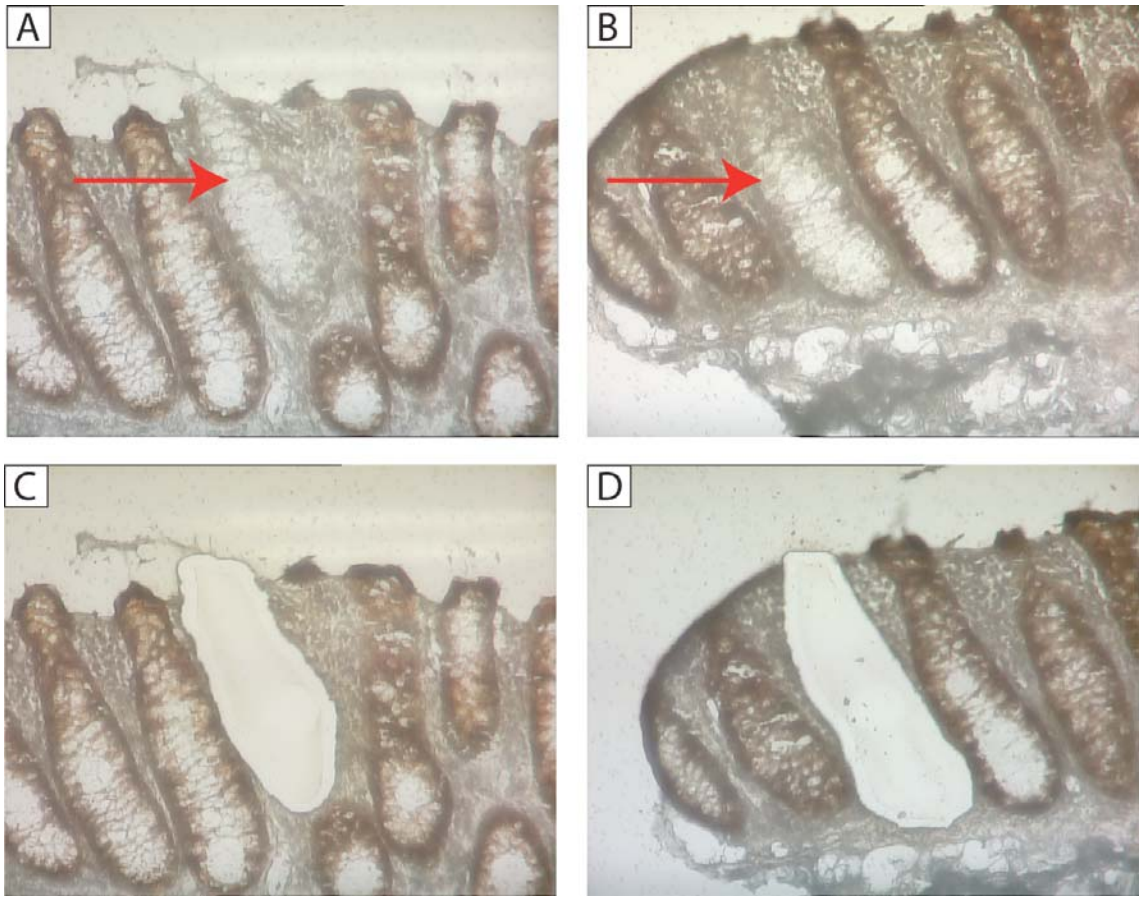

Figure 4.1 LCMD of whole CCO-deficient crypts for MS analysis
CCO-histochemistry on serial sections of fresh frozen human colon; no SDH counter-stain was performed. **(A-B):** Lack of DAB staining identifies CCO-deficient crypts (highlighted by red arrow). **(C-D):** Post LCMD of corresponding crypts shown in **(A)** and **(B)**. All images are at 200x magnification.

Tubes containing 10 and 5 whole crypts underwent MALDI-MS analysis as per the protocol detailed in Methods 2.9. Unfortunately, no peptides could be detected on digests of multiple crypts, so no further studies were carried out since the technique would not be sensitive enough to analyse CCO-deficient clones within individual crypts.

4.3 RT-PCR for lineage markers

Another technique to phenotype cells is to use mRNA transcription as a proxy marker for protein expression. RNA extraction and RT-PCR protocols were designed, optimised and performed on whole crypts that had been stained for CCO-activity (see Methods 2.8). Extracting RNA from frozen tissue presents many more problems than that for DNA, since exposure to the aqueous phase activates endogenous RNAses that rapidly degrade intracellular RNA. The CCO-histochemistry protocol was therefore shortened, in order to limit exposure to the aqueous phase, by omitting the SDH counter-stain that turns CCO-deficient cells blue. Nested PCR primers specific for the mRNA of the following lineage specific markers were designed: the brush border proteins villin and CD10 (enterocytes - absorptive lineage); Mucin glycoprotein 2 (MUC-2) (goblet cells – secretory lineage); Chromogranin-A (enteroendocrine cells – secretory lineage). In order to provide control complementary DNA (cDNA) with which to optimise the PCR protocols, a small piece of fresh frozen human colonic mucosa was Trizol® digested and cDNA synthesised as detailed in Methods 2.8.

Figure 4.2A demonstrates an agarose gel from an RT-PCR for villin, MUC-2, CD10 and chromogranin A mRNA using Trizol® digested human colon cDNA of varying dilutions. Clear, strong bands of the appropriate size were seen for villin, MUC-2 and chromogranin-A (Figure 4.2A). The RT-PCR products were then sequenced and found to be specific for the respective target mRNA sequences. To test the sensitivity of the technique for individual crypts, whole CCO-proficient and CCO-deficient crypts were micro-dissected, followed by RNA extraction and RT-PCR. Unfortunately mRNA for villin, MUC-2 and chromogranin-A could not be consistently amplified from individual colon crypts (Figure 4.2B), therefore the technique was not considered sensitive or reliable enough for determining cell lineages within CCO-deficient clones.

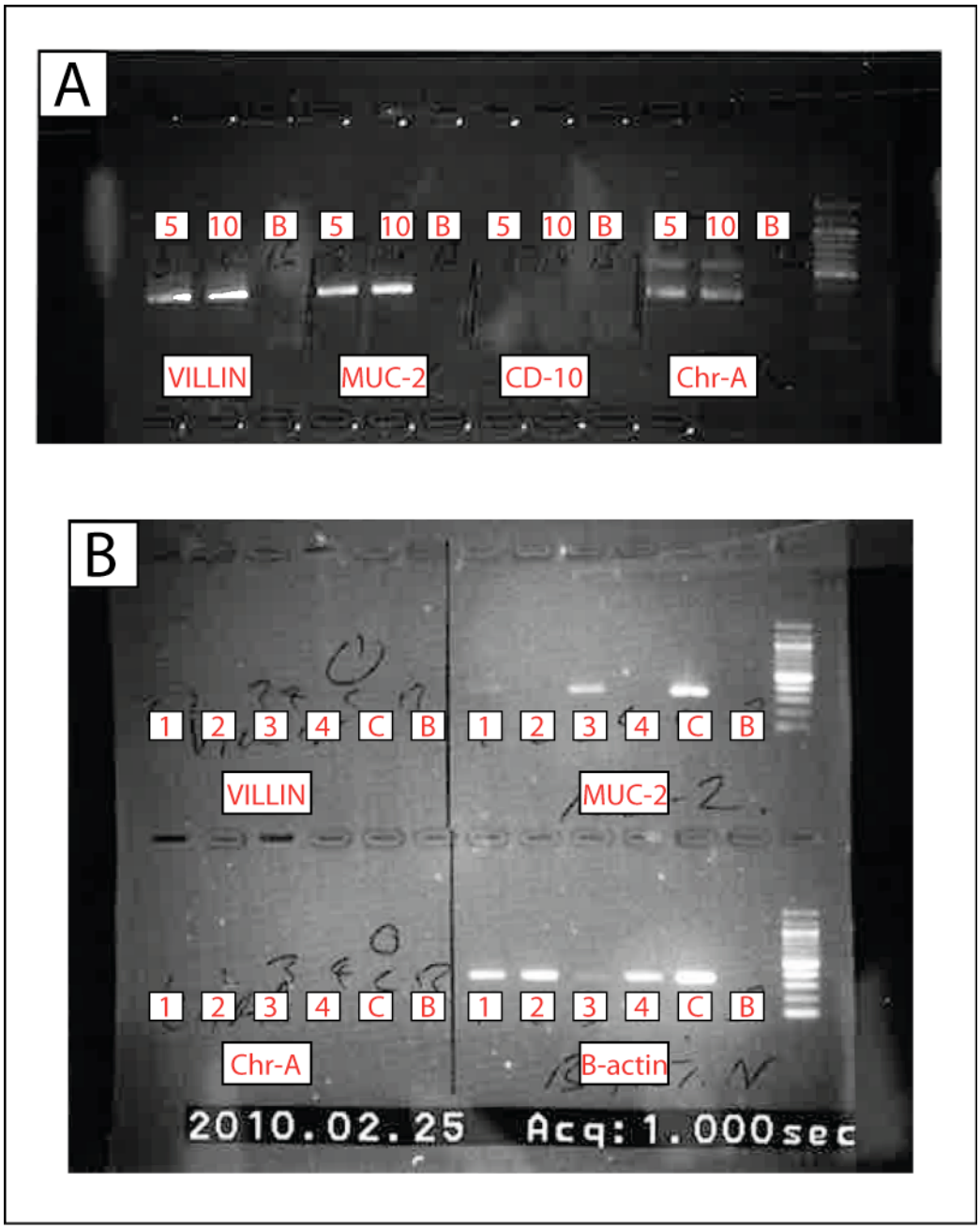

Figure 4.2 RT-PCR results for control cDNA and individual colon crypts
(A): Agarose gel for RT-PCR products using nested primers specific for the mRNA of villin, MUC-2, CD10 and chromogranin A (Chr-A) on control cDNA. The 3 lanes for each primer set were 5x diluted control cDNA (5), 10x diluted control cDNA (10) and an RT negative control (B); 100bp ladder is the far right lane. **(B):** Agarose gel for nested RT-PCR on individual crypts; lanes 1 and 2 were CCO-proficient crypts; lanes 3 and 4 CCO-deficient crypts; lane 5 was control cDNA and lane 6 an RT negative control. Positive bands seen for MUC-2 with crypts 1 and 3 only; β-Actin was used as an internal control and was positive for all crypts.

4.4 Immunohistochemistry and immunofluorescence of CCO-deficient clones for lineage markers

Most cells within the colonic crypt are derived either from the absorptive (enterocytes) or secretory lineages (goblet cells and enteroendocrine cells); Paneth cells are only occasionally found in the right colon. Therefore immunohistochemical and immunofluorescent staining protocols were designed to combine identification of CCO-deficient clones within crypts, with lineage labelling for the absorptive and secretory lineages (see Methods 2.6) in order to demonstrate multi-lineage potential.

4.4.1 Immunofluorescence for MUC-2 reliably identifies goblet cells of the secretory lineage in the human colon crypt

The predominant secretory cell type of colon crypts is the mucin glycoprotein producing goblet cell. Therefore immunofluorescence (IF) using an antibody against Mucin-2 (MUC-2), which is specific to intestinal epithelial goblet cells, was used as a proxy marker for the secretory cell lineage in crypts (See Methods 2.6.2).

Figure 4.3 illustrates MUC-2 IF performed on normal human colon crypts using both FFPE and fresh frozen tissue, as detailed in Methods 2.6.2; staining is specific for goblet cells, which are easily identified by their intra-cytoplasmic vacuoles.

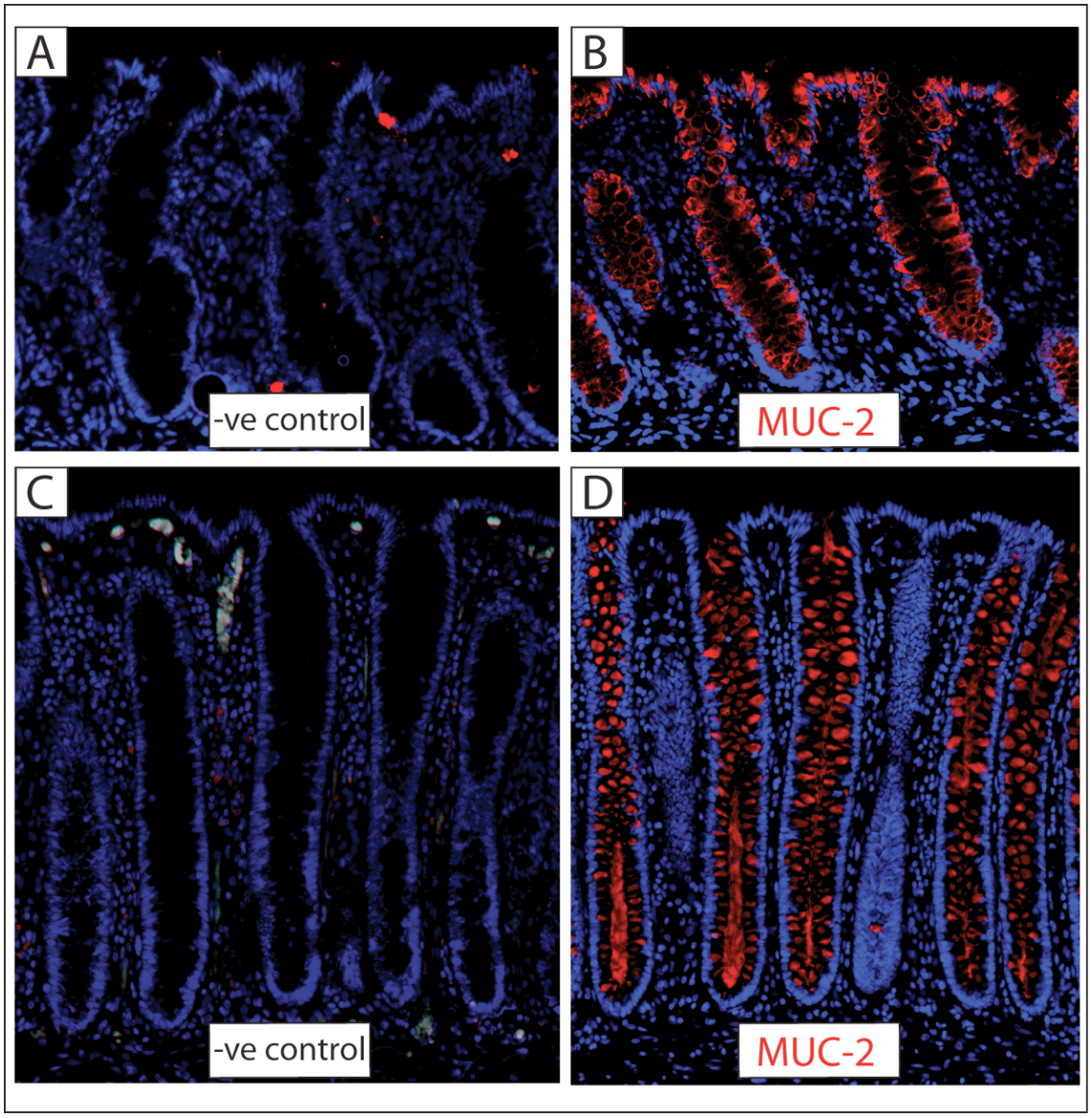

Figure 4.3 Mucin-2 (MUC-2) staining of normal human colon crypts
(A-B): MUC-2 IF (Cy3/555nm) on fresh frozen human colon; negative control on the left **(A)**, MUC-2 IF is specific for colon crypt goblet cells **(B)**. **(C-D):** MUC-2 IF (Cy3/555nm) on FFPE human colon; negative control on the left **(C)**, staining is specific for crypt goblet cells **(D)**. DAPI nuclear counter-stain in blue for all panels and all images are at 100x magnification.

4.4.2 Immunofluorescence for carbonic anhydrase II (CA-II) reliably marks cells of the absorptive lineage in human intestinal crypts

Production of the protein Carbonic Anhydrase II (CA-II) has been previously shown to be specific for progenitor cells committed to the absorptive lineage and differentiated enterocytes of the intestine [177, 178]. In order to validate this antibody, immunofluorescence was performed on human small intestine using CD10, known to specifically mark the brush border of small bowel enterocytes, as a positive control (Figure 4.4A-B). CA-II demonstrated a similar staining pattern to CD10 in the human small bowel (Figure 4.4C-D), specifically marking the differentiated enterocytes of the villi. In the human colonic crypt CA-II staining was again specific for non-goblet cells (Figure 4.5A-D), marking cells from the mid-crypt consistent with when progenitor cells become committed to an absorptive lineage, as validated by previous authors [177, 178]. Therefore CA-II staining was used as a proxy marker for the absorptive cell lineage in colon crypts.

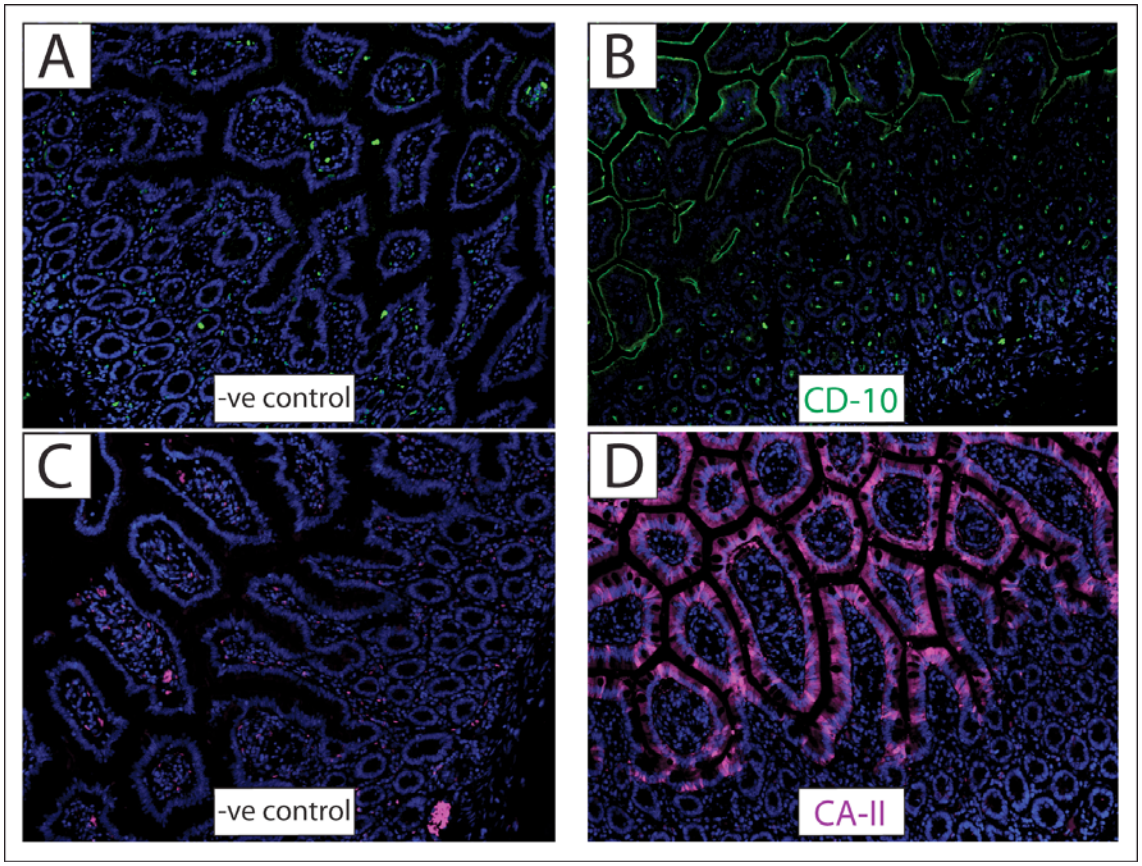

Figure 4.4 Carbonic Anhydrase II (CA-II) staining of normal human small bowel crypts
(A-B): CD10 IF (green/488nm) in human FFPE small intestine; negative control on the left **(A)**, staining is specific to the enterocyte brush border **(B)**. **(C-D):** CA-II IF (Cy5/647nm) of human FFPE small intestine; negative control is on the left **(C)**, in **(D)** CA-II is seen to specifically stain the differentiated enterocytes of the villi, with goblet cells unstained. DAPI nuclear counter-stain in blue for all panels and all images are at 100x magnification.

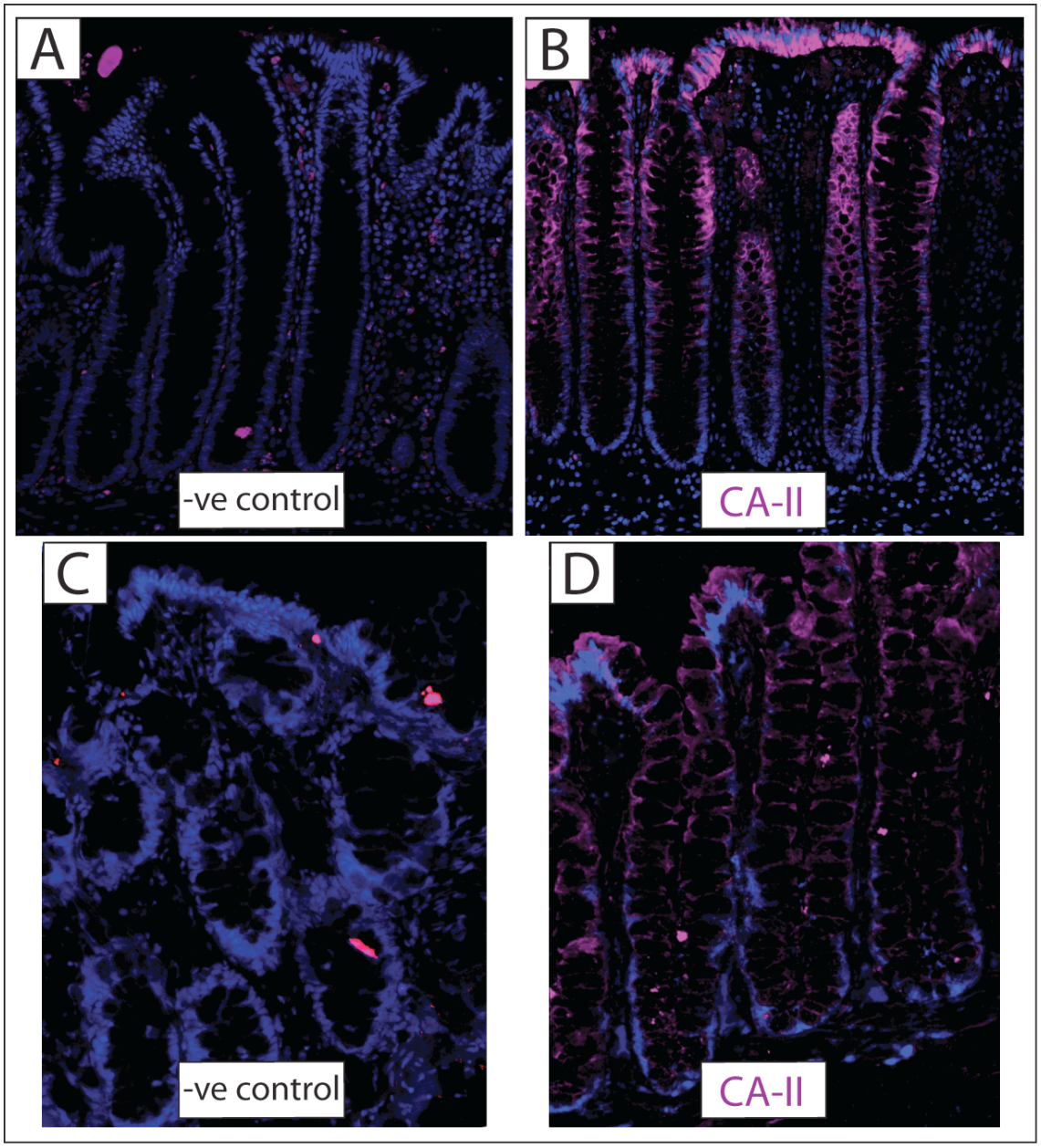

Figure 4.5 CA-II staining of FFPE and fresh frozen normal human colon crypts **(A-B):** CA-II IF (Cy5/647nm) of human FFPE colon; negative control on left **(A)**, CA-II staining is specific to the non-goblet cells of the mid-crypt, upper crypt and mucosal surface **(B)**. **(C-D):** CA-II IF (Cy5/647nm) of fresh frozen human colon; negative control on the left **(C)**, again, CA-II specifically stains the non-goblet cell lineage **(D)**, with increasing intensity in the mid and upper crypt. DAPI nuclear counter-stain is in blue for all panels and all images are at 100x magnification.

4.4.3 Immunofluorescence demonstrates multi-lineage potential of CCO-deficient crypts and CCO-deficient clones

Immunohistochemistry for cytochrome *c* oxidase subunit I, encoded for by the mitochondrial genome, has been shown to reliably detect CCO-deficient crypts and correspond with the absence of CCO enzyme activity [76]. Therefore, triple immunofluorescence for CCO-subunit I, MUC-2 and CA-II was performed on serial sections of normal FFPE tissue, in order to identify cells of the secretory and absorptive lineage in wholly and partially CCO-deficient crypts (see Methods 1.6.2). Figures 4.6-4.8 demonstrate single and patches of CCO-deficient crypts that stained positively for both MUC-2 and CA-II, confirming that CCO-deficient crypts contain cells of both lineages and appear otherwise identical to adjacent wild-type, CCO-proficient crypts.

In addition, as with enzyme histochemistry, CCO-deficient areas were identified within crypts (Figures 4.9A and 4.10A). These CCO-deficient cell populations clearly stained for cells of both the secretory and absorptive lineage (Figure 4.9B-D and Figure 4.10B-D), with no difference in staining patterns observed compared to the CCO-proficient areas within the partially-deficient crypts or adjacent wild-type crypts.

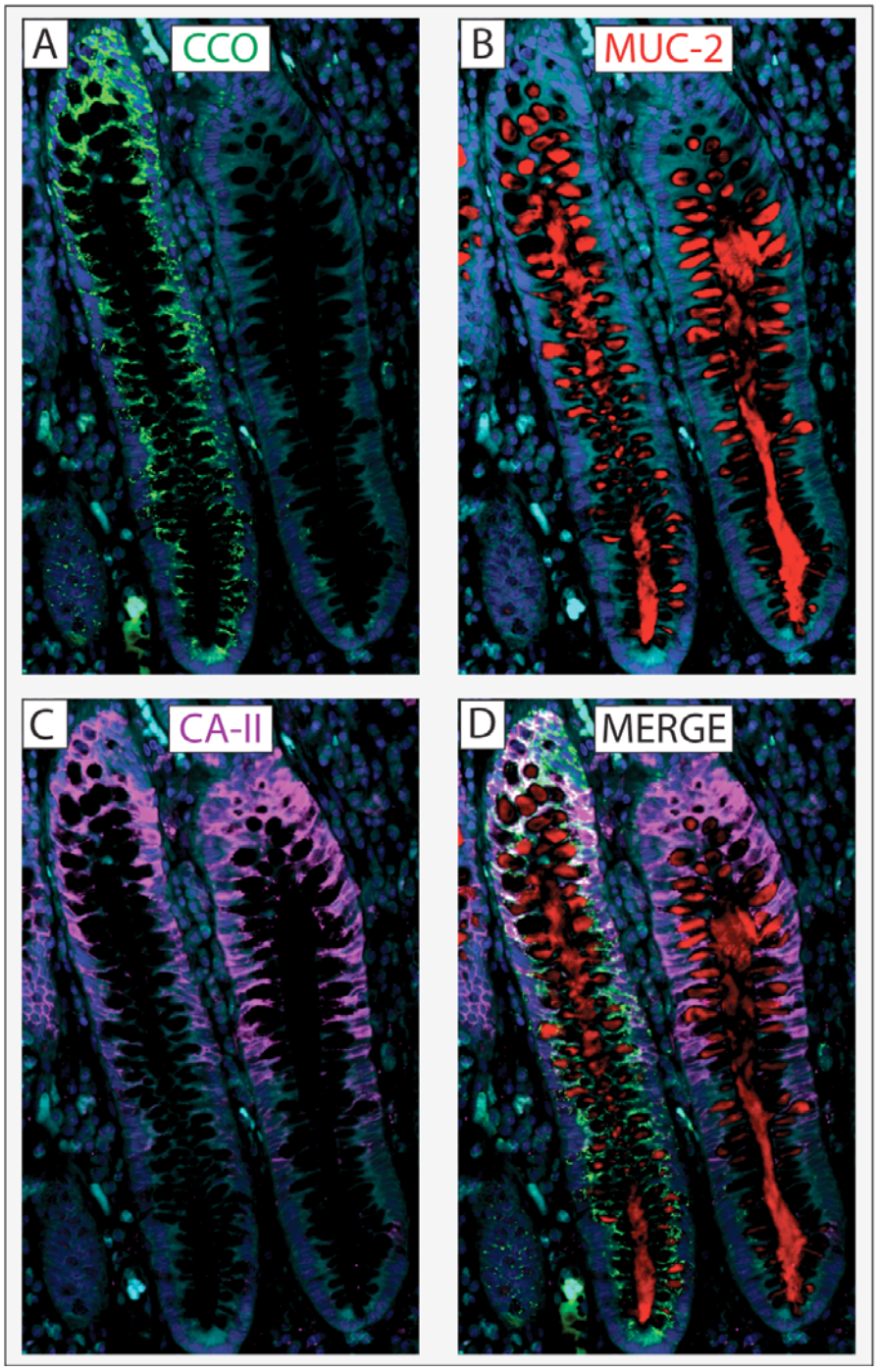

Figure 4.6 CCO-deficient crypts contain cells of both the secretory and absorptive lineages

Sections shown are from FFPE normal human colon. Triple IF for CCO (green/488nm), MUC-2 (Cy3/555nm) and CA-II (Cy5/647nm) has been performed on the same section, and demonstrates a CCO-deficient crypt (right) **(A)** containing cells of both the secretory **(B)** and absorptive **(C)** lineage respectively, adjacent to a CCO-proficient crypt (left); merged images shown in **(D)**. For all panels DAPI nuclear counter-stain (blue) and auto-fluorescence in the Aqua channel have been used to aid morphological definition of cells. Images are at 200x magnification.

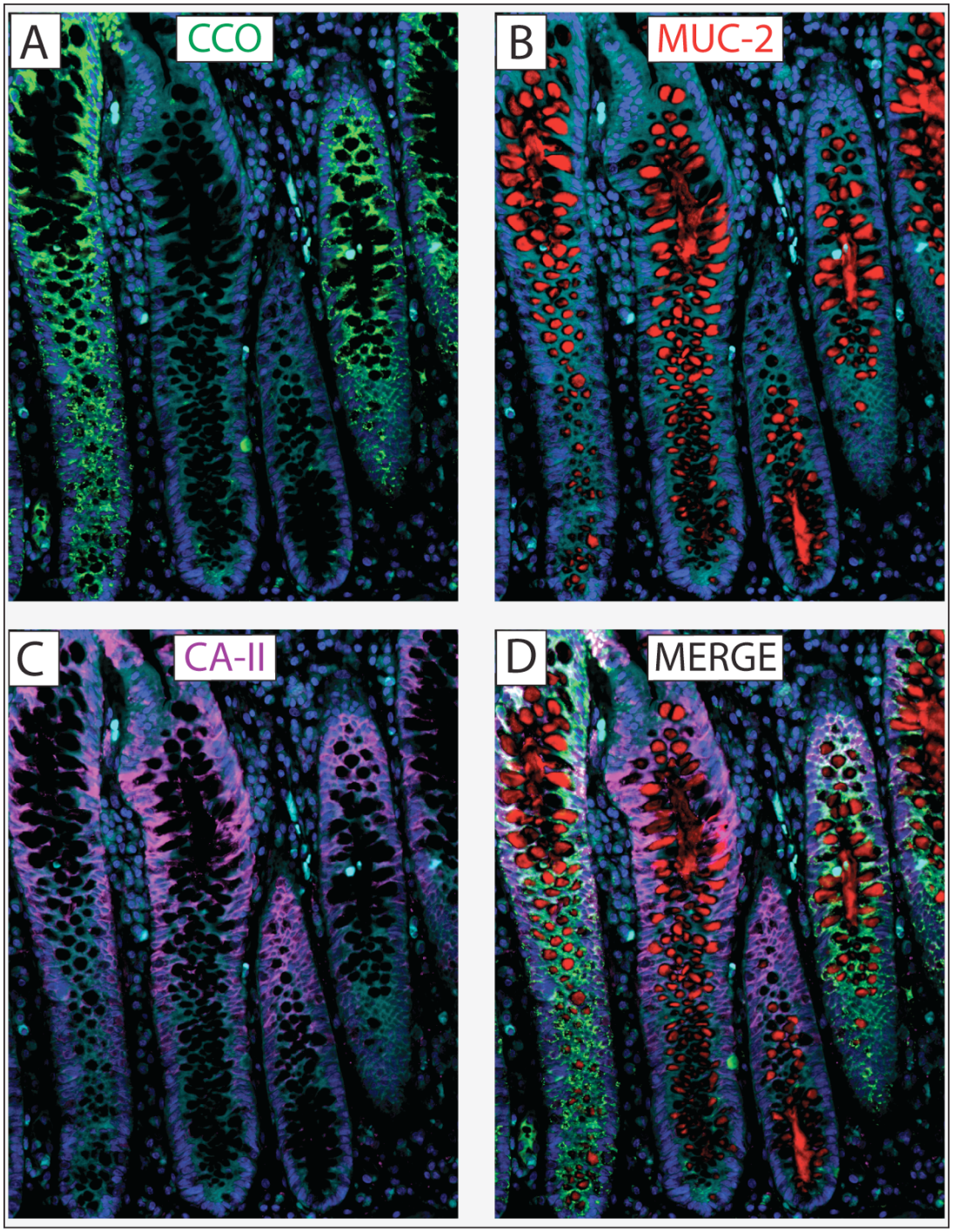

Figure 4.7 Patches of CCO-deficient crypts contain cells of both the secretory and absorptive lineages (1)
Sections shown are from FFPE normal human colon. Triple IF for CCO (green/488nm), MUC-2 (Cy3/555nm) and CA-II (Cy5/647nm) has been performed on the same section, and demonstrates a patch of two CCO-deficient crypts **(A)** containing cells of both the secretory **(B)** and absorptive **(C)** lineages; merged images shown in **(D)**. For all panels DAPI nuclear counter-stain (blue) and auto-fluorescence in the Aqua channel have been used to aid morphological definition of cells. Images are at 100x magnification.

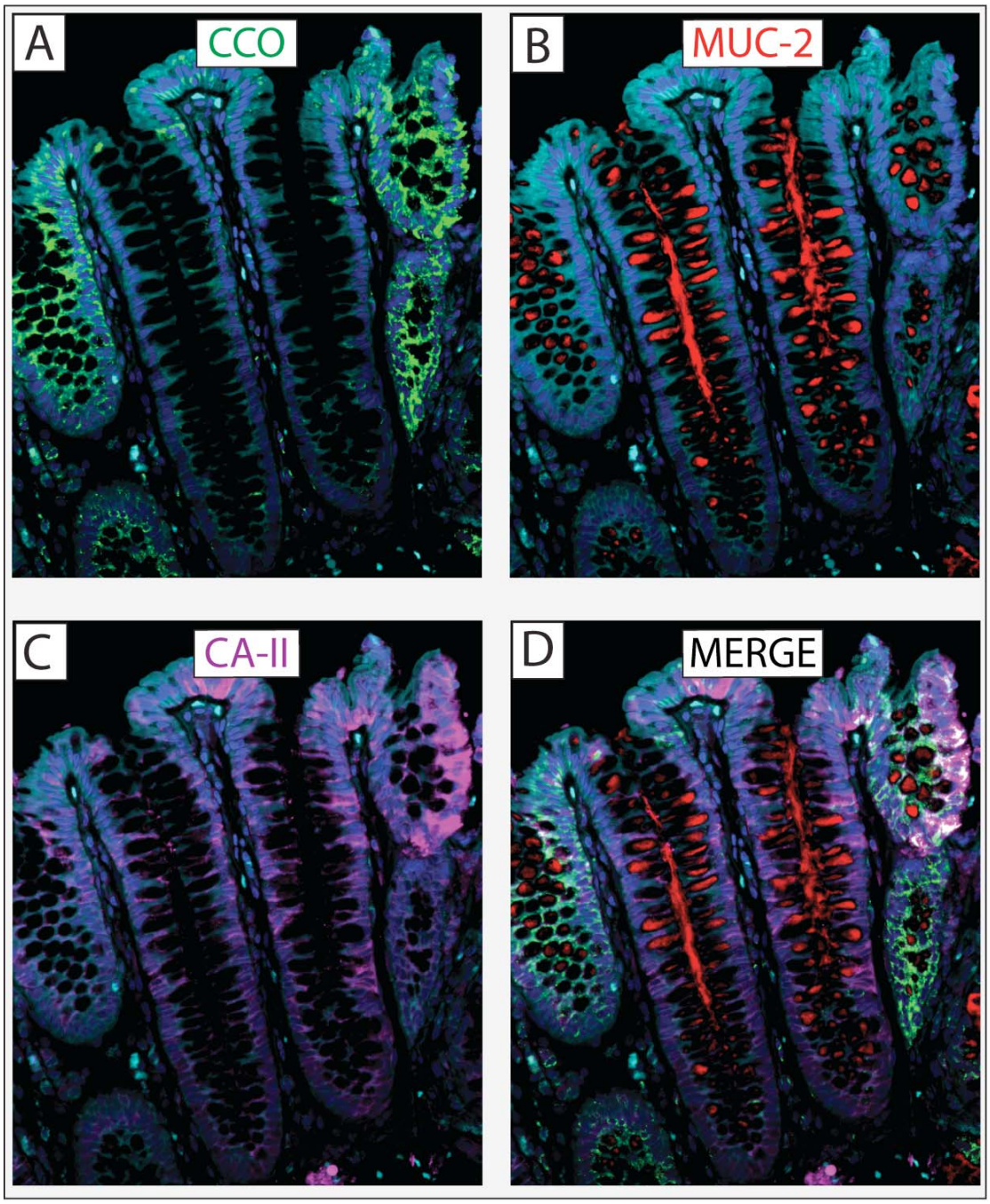

Figure 4.8 Patches of CCO-deficient crypts contain cells of both the secretory and absorptive lineages (2)
Sections shown are from FFPE normal human colon. Triple IF for CCO (green/488nm), MUC-2 (Cy3/555nm) and CA-II (Cy5/647nm) has been performed on the same section, and demonstrates a patch of two CCO-deficient crypts **(A)** containing cells of both the secretory **(B)** and absorptive **(C)** lineages; merged images shown in **(D)**. For all panels DAPI nuclear counter-stain (blue) and auto-fluorescence in the Aqua channel have been used to aid morphological definition of cells. Images are at 100x magnification.

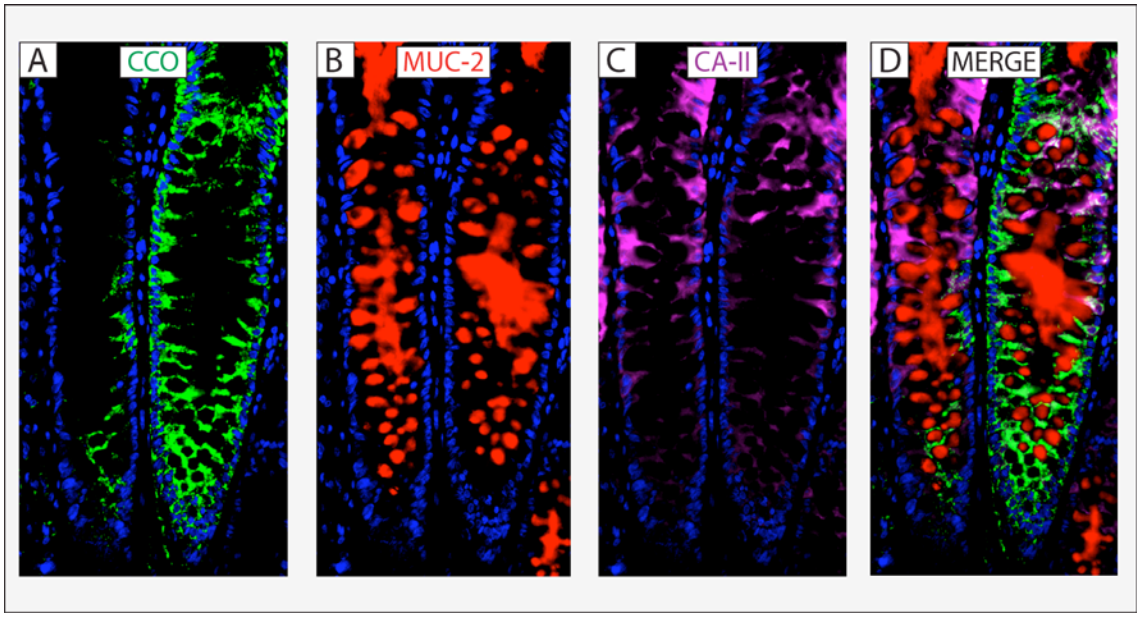

Figure 4.9 CCO-deficient cell populations within crypts contain both secretory and absorptive lineages (1)
Sections shown are from FFPE normal human colon. Triple IF for **(A)** CCO (green/488nm), **(B)** MUC-2 (Cy3/555nm) and **(C)** CA-II (Cy5/647nm) has been performed on the same section; merged image is shown in **(D)**. **(A)** CCO-deficient cell population is seen within a partially-deficient crypt on the left, adjacent to a wild-type, CCO-proficient crypt (right); IF demonstrates cells of the secretory **(B)** and absorptive lineage **(C)** within the CCO-deficient area of the crypt. For all panels DAPI nuclear counter-stain is in blue and images are at 200x magnification.

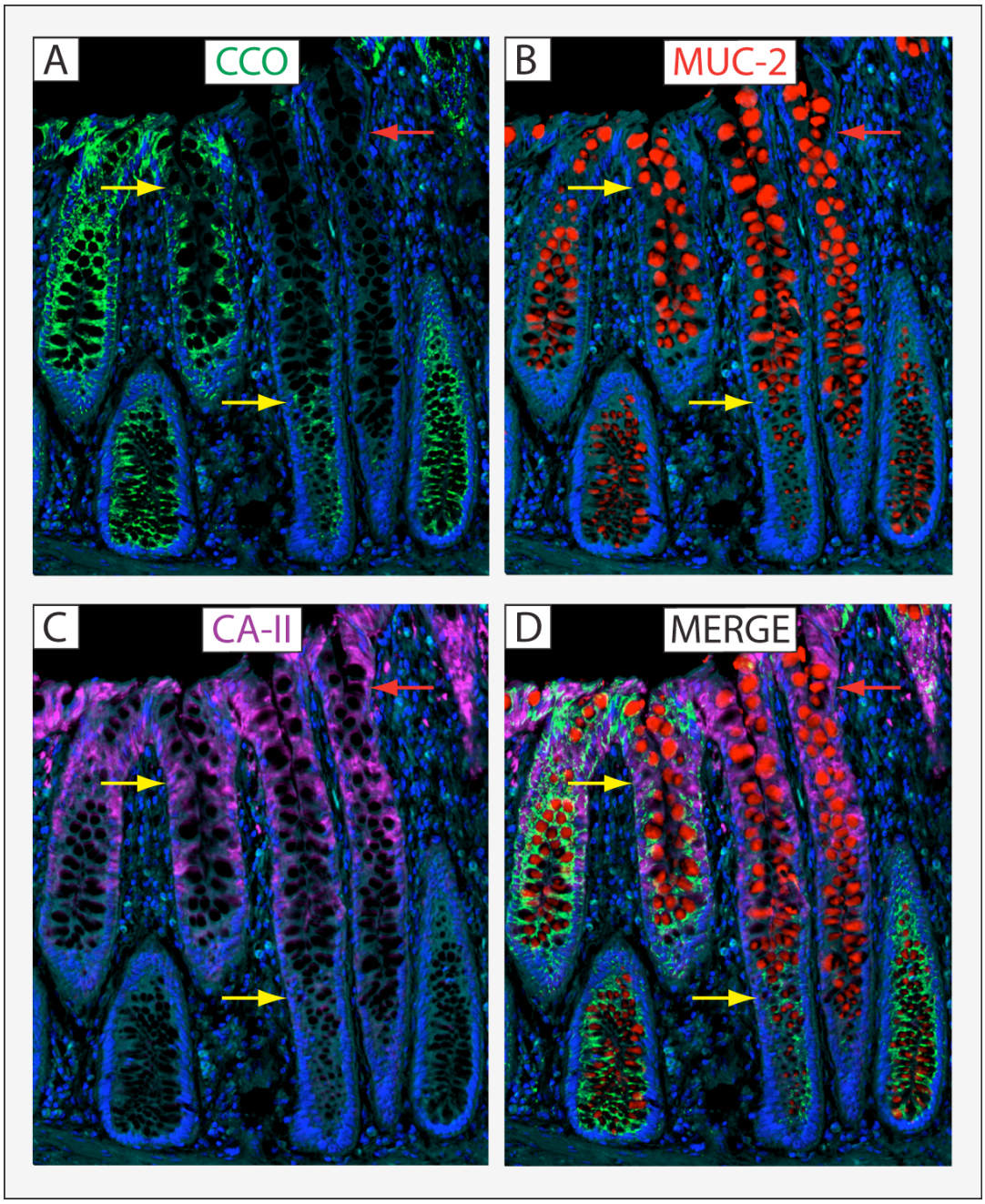

Figure 4.10 CCO-deficient cell populations within crypts contain both secretory and absorptive lineages (2)
Sections shown are from FFPE normal human colon. Triple IF for **(A)** CCO (green/488nm), **(B)** MUC-2 (Cy3/555nm) and **(C)** CA-II (Cy5/647nm) has been performed on the same section of FFPE normal human colon; merged image is shown in **(D)**. **(A)** CCO-deficient cell populations are seen within partially-deficient (yellow arrows) and wholly-deficient crypts (red arrow) adjacent to wild-type, CCO-proficient crypts; IF demonstrates cells of the secretory **(B)** and absorptive lineage **(C)** within the CCO-deficient populations. For all panels DAPI nuclear counter-stain (blue) and auto-fluorescence in the Aqua channel have been used to aid morphological definition of cells. All images are at 100x magnification.

Traditionally, immunohistochemistry and immunofluorescence is best performed on FFPE tissue, as the anatomy is better preserved and thus the staining of much better quality; however, enzyme histochemistry for CCO-activity to identify CCO-deficient clones within crypts can only be done on fresh frozen tissue. Thus, alternate serial sections of normal human frozen colon were stained for CCO-histochemistry in order to identify partially CCO-deficient crypts containing putative CCO-deficient stem cell clones (Figure 4.11A and Figure 4.12A and 4.12C). Once a partial crypt was identified, the intervening sections were stained using dual immunofluorescence for MUC-2 to mark the secretory cell lineage and CA-II to mark the absorptive cell lineage. CCO-deficient clones again demonstrated staining for both cell lineages (Figure 4.11B and Figure 4.12B), as seen with FFPE tissue.

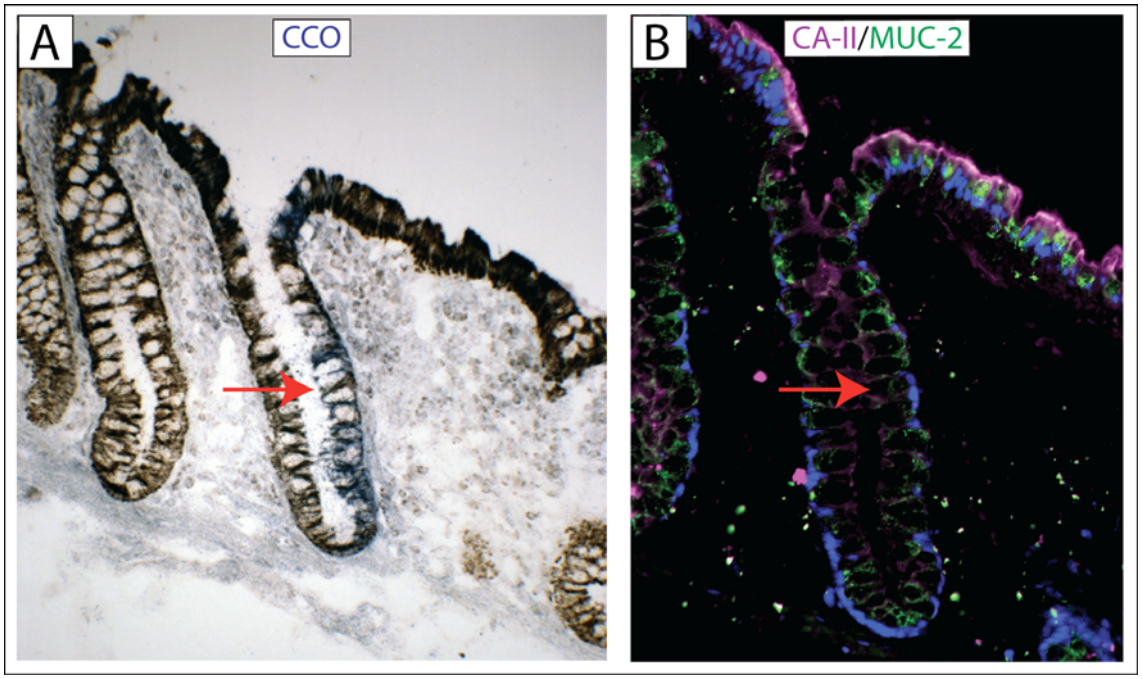

Figure 4.11 Combined CCO-histochemistry and immunofluorescence confirms CCO-deficient populations within human colon crypts contain cells of both the secretory and absorptive lineages (1)
(A): CCO-histochemistry of on-edge serial sections of fresh frozen human colon identifies a crypt (highlighted by red arrows) containing a CCO-deficient (blue) cell population; **(B):** Double IF for MUC-2 (green/488nm) and CA-II (Cy5/647nm) on the next serial section: cells of the absorptive and secretory lineage are demonstrated within the CCO-deficient area. DAPI nuclear counter-stain is in blue and all images are at 200x magnification.

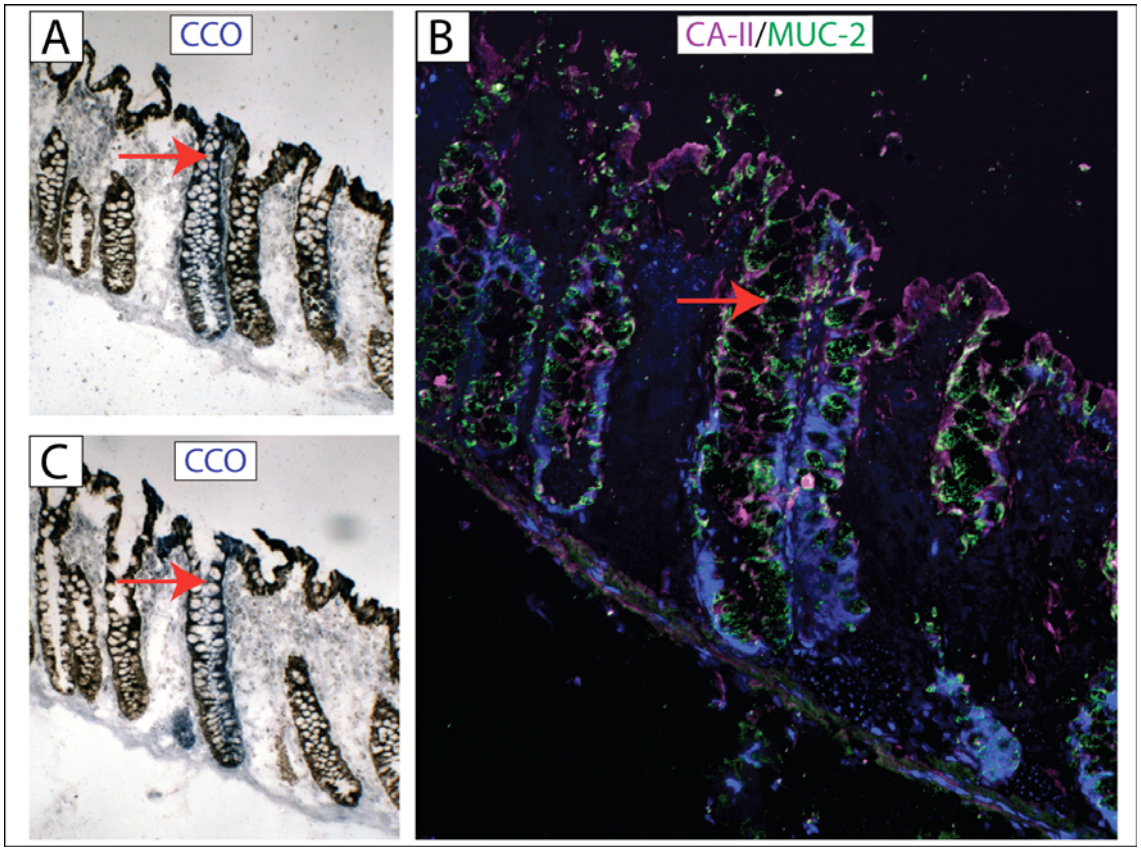

Figure 4.12 Combined CCO-histochemistry and immunofluorescence confirms CCO-deficient populations within human colon crypts contain cells of both the secretory and absorptive lineages (2)
(A-C): Serial sections of fresh frozen human colon. **(A) & (C):** CCO-histochemistry identifies a crypt containing a CCO-deficient (blue) cell population (highlighted by red arrows) on serial sections. **(B):** Double IF for MUC-2 (green/488nm) and CA-II (Cy5/647nm) on the intervening serial section between **(A)** and **(C)** demonstrates cells of the absorptive and secretory lineage within the CCO-deficient clone. DAPI nuclear counter-stain is in blue and all images are at 100x magnification.

4.5 *In-situ* hybridisation for the stem cell marker *LGR5* identifies basal stem cells within CCO-deficient and wild-type crypts

Expression of the Wnt target gene *LGR5* has been shown to be a reliable marker of colonic stem cells in mice [19]. In order to demonstrate that CCO-deficient crypts contain stem cells at the base and are no different to CCO-proficient, wild-type crypts, *in situ* hybridisation (ISH) for *LGR5* mRNA was performed on serial sections of FFPE normal human colon (see Methods 2.7), with immunohistochemistry for cytochrome *c* oxidase subunit I on intervening alternate serial sections. ISH does not work well on fresh frozen tissue due to the activation of endogenous RNAses and disruption of the anatomy as a result of the multiple digestion steps. In Figure 4.13, both wild-type and a CCO-deficient crypt can be seen to exhibit similar signal for *LGR5* mRNA that is specific to the crypt base and the area of the putative stem cell niche.

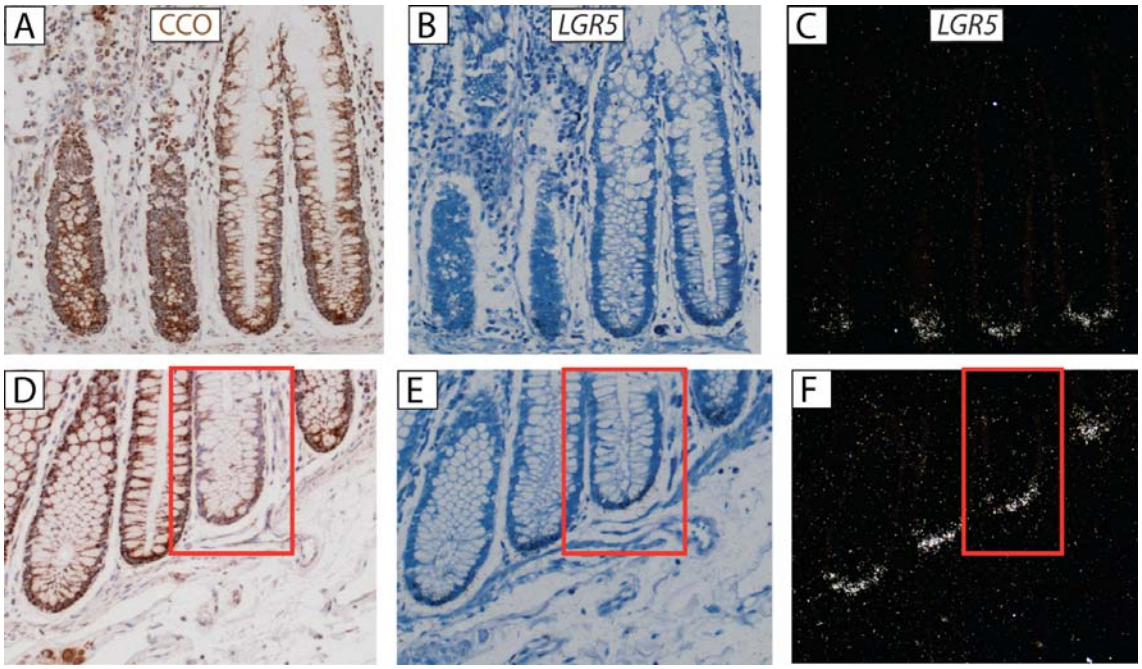

Figure 4.13 CCO-deficient and CCO-proficient human colon crypts contain basal stem cells
(A-C): CCO-immunohistochemistry showing four CCO-proficient, wild-type crypts (A); S^{35} ISH for *LGR5* mRNA on next serial section is specific to the base of the crypts, in the area of the putative stem cell niche, bright-field (B) and dark-field images (C) shown. (D-E): Immunohistochemistry for CCO-subunit I with a CCO-deficient crypt highlighted by the red box, surrounded by wild-type crypts (D); ISH for *LGR5* mRNA is again specific to the basal stem cell niche for all crypts (E-F). All images are at 200x magnification.

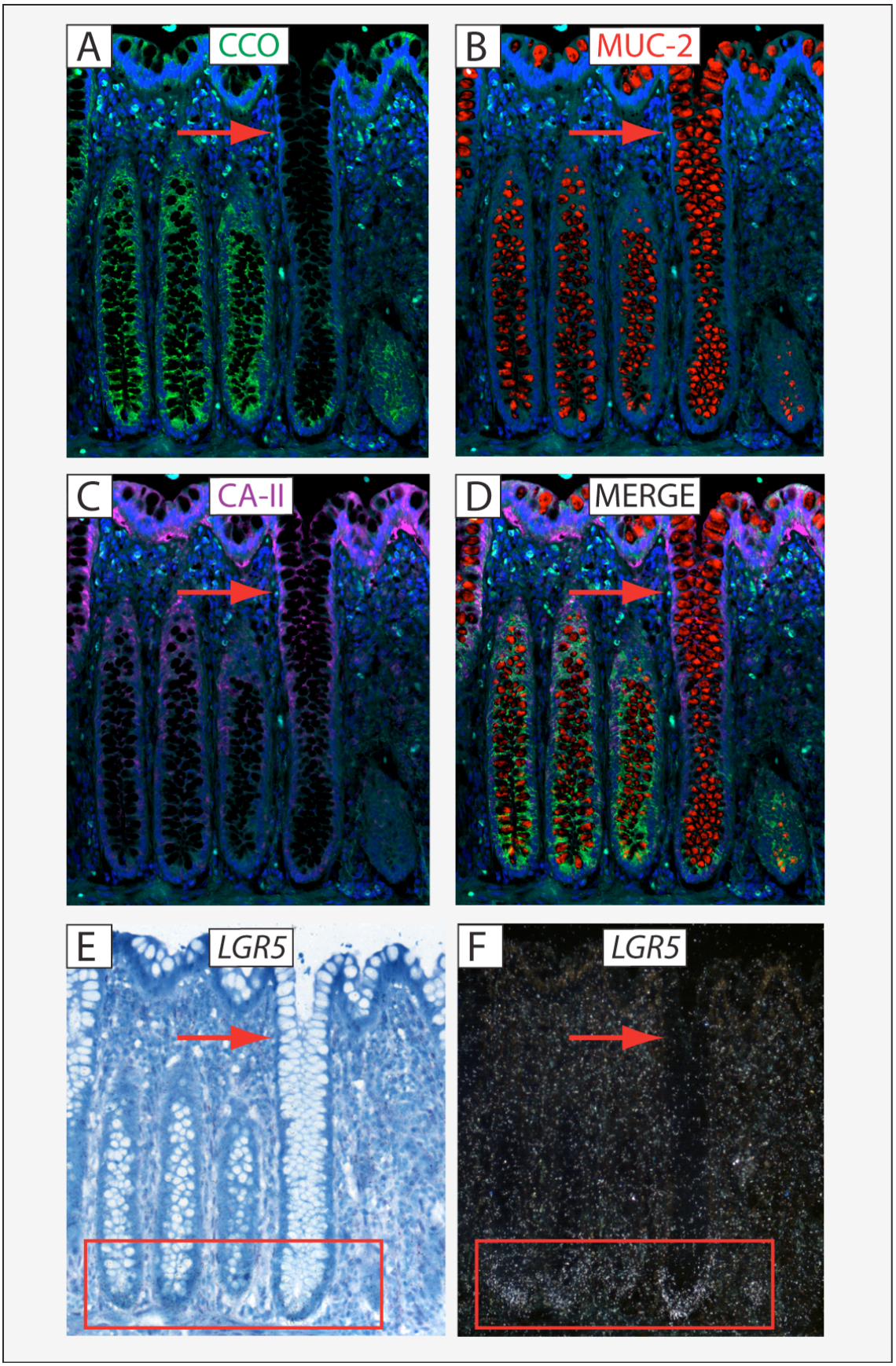
A
CCO
B
MUC-2
C
CA-II
D
MERGE
E
LGR5
F
LGR5

Figure 4.14 CCO-deficient cell populations within crypts are multi-potential and stain for stem cells at the crypt base
(A-D): Triple IF for **(A)** CCO-subunit I (green/488nm), **(B)** MUC-2 (Cy3/555nm) and **(C)** CA-II (Cy5/647nm) has been performed on the same section; merged image is shown in **(D)**. **(A)** A CCO-deficient cell population seen within a partially deficient (red arrows) crypt adjacent to wild-type, CCO-proficient crypts; IF demonstrates cells of the secretory **(B)** and absorptive lineage **(C)** within the CCO-deficient population. For all panels DAPI nuclear counter-stain (blue) and auto-fluorescence in the Aqua channel have been used to aid morphological definition of cells. **(E-F):** ISH for *LGR5* mRNA on the next serial section marks putative stem cells at the crypt base; bright-field **(E)** and dark-field **(F)** images shown. All images are at 100x magnification.

In order to identify partially CCO-deficient crypts containing multi-potential putative stem cell clones, triple IF for MUC-2 and CA-II was performed on FFPE normal human colon as previously (Figure 4.14A-D). ISH for *LGR5* mRNA performed on intervening serial sections marked the stem cell population at the base of CCO-deficient clones, with no difference observed between the remaining CCO-proficient cells within the crypt or adjacent wild-type crypts (Figure 4.14E-F). The background ISH staining, representing low-level *LGR5* expression and non-specific binding of the riboprobe, is high in Figure 4.14, however a high intensity signal specific to the crypt base can clearly be seen in both the bright and dark-field images.

4.6 Discussion

Previous studies have shown that all cells within CCO-deficient crypts harbour identical, clonal somatic point mutations in their mtDNA as a result of their accumulation in stem cells [76], and that patches of adjacent CCO-deficient crypts contain identical mutations in their mtDNA [77]. The minimal odds of this occurring by chance are $\geq 2.48 \times 10^9$:1, thus this can only be the consequence of crypt fission. In Chapter 3, it has been shown that CCO-deficient cell populations within crypts represent the progeny of a single stem cell lineage, since all CCO-deficient cells harbour clonal mutations in their mtDNA, not present in the remaining CCO-proficient, wild-type cells of the same crypt. However, are the CCO-deficient cell populations behaving normally? More

recent studies demonstrating the clonality of CCO-deficient glands in the human small intestine and human stomach [78, 79] have shown that CCO-deficient glands here do contain all the expected differentiated cell types. Noteboom et al [82] have shown there to be a small but significant decrease in the proliferative index in human colonic CCO-deficient crypts versus wild-type crypts (average decease 16% as measured by Ki-67) and increase in the apoptotic frequency in colonic CCO-deficient cells (0.33%) versus CCO-proficient cells (0.29%). However, there is now direct evidence to show that significant respiratory chain defects are neutral and do not adversely affect epithelial stem cells: Baris et al [179] selectively ablated the respiratory chain in epidermal progenitor stem cells and did not find any alteration in differentiated cell types, proliferation or apoptotic indices, despite there being a complete absence of respiratory chain complexes.

The results of the IF studies presented in Figures 4.6-4.8 demonstrate that individual and patches of wholly CCO-deficient crypts contain both secretory and absorptive lineages, and that staining patterns are no different to that of adjacent wild-type, CCO-proficient crypts. Again, when studying CCO-deficient cell populations within crypts, using triple IF and IF combined with CCO-histochemistry, both cell lineages are clearly seen, with similar staining patterns to that of the wild-type cells within the crypt (Figures 4.9-4.12). Thus, CCO-deficiency does not appear to alter the cell population within human colon crypts, and the progeny of distinct CCO-deficient clones within colon crypts have multi-lineage potential.

The Wnt target gene *LGR5* has been shown to be a robust intestinal stem cell marker in mice in both the small and large intestine [19] and is analogous in humans. *In–situ* hybridisation (ISH) for *LRG5* mRNA was specific for the base of the crypts in both CCO-deficient and CCO-proficient crypts (Figure 4.13), supporting this as the location of the stem cell niche in human colon crypts. Unfortunately, no small bowel tissue was available for analysis. Furthermore, *LGR5* ISH also marked putative stem cells at the base of CCO-deficient clones within crypts (Figure 4.14), with no difference in signal

observed between the two populations, again suggesting that CCO-deficiency does not alter the stem cell architecture. There is evidence that the intestinal stem cell niche is a heterogeneous population: Montogomery et al [29] identified *mTert* expression as marking multipotent, long-lived stem cells, distinct to the Lgr5+ population, in both the mouse small bowel and colon, thus further work using a panel of putative stem cell markers on CCO-deficient stem cell clones in both the small and large bowel might give insights into whether distinct sub-populations of stem cells exist within the human intestinal niche.

In summary, the results from this chapter demonstrate that, not only are human colonic CCO-deficient clones the progeny of a single stem cell lineage within the crypt, they contain both committed secretory and absorptive cells, thus proving their multi-lineage potential and fulfilling the criteria for identifying these as stem cell clones. The niche, as identified by ISH for the putative intestinal stem cell marker *LGR5*, is at the base of human colonic crypts, consistent with observations from mouse lineage tracing studies [19, 29].

Chapter 5. Clonal expansion of normal human colonic crypts

5.1 Introduction and aims

Stem cell dynamics in human epithelial tissues, such as the colonic crypt, remain uncertain, due to the inherent limitations of using human tissue. However, methylation patterns of non-expressed genes at CpG island loci offer a serendipitous means to study dynamics and infer rates of clonal expansion in human tissues [18, 45, 168] (see Introduction 1.1.3). Methylation at these CpG loci is somatically inherited, and the CpG islands are initially un-methylated in the zygote but acquire *de-novo* methylation stochastically at mitosis. Therefore, comparison of the methylation patterns between two cells should reveal their clonal relation: cells with a recent common ancestor should have similar methylation patterns whereas unrelated, or more distantly related cells, are unlikely to share similar patterns of methylation. This makes methylation patterns an attractive means to study the ancestry and dynamics of cells within the human body, without the need for invasive labelling techniques.

The aim of this project was to study the methylation patterns at these CpG islands of clonal, CCO-deficient cell populations in the colonic mucosa, in order to investigate the dynamics of the colonic stem cell and rates of clonal expansion. It was hypothesised that (i) analysis of the methylation patterns of CCO-deficient clones within crypts would enable the dynamics of niche succession to be estimated, and (ii) by comparing methylation patterns of adjacent, wholly CCO-deficient crypts, previously shown to be clonally derived by fission [77], with unrelated CCO-proficient crypts, the dynamics of clonal expansion in the normal epithelium could be investigated.

Tissue was collected from 9 patients who were undergoing surgery for either a colorectal polyp or cancer (see Appendix 9.4), and morphologically normal tissue was obtained from outside the resection margins of the polyp or cancer.

5.2 Validation of methylation methods in human colon crypts

The CpG loci studied, within non-expressed genes, were the same as those used by previous studies to analyse human colon crypt stem cell dynamics: (i) *CSX*: Cardiac specific homeobox, expressed specifically during the differentiation of cardiac myocytes; (ii) *MYOD1*: Myoblast determination protein-1, specifically expressed during myoblast differentiation; (iii) *BGN*: Biglycan, an X-chromosome cellular matrix proteglycan expressed in connective tissue. The *CSX*, *MYOD1* and *BGN* CpG islands have eight, five and ten CpG sites respectively, and are located in the coding sequence of these non-expressed genes. PCR amplification of bisulphite-treated DNA using methylation specific primers followed by cloning of individual PCR strands allows the methylation status of each gene-specific CpG island for a defined cell population – an individual crypt or clones within crypts - to be sampled (see Methods 2.10 & Figure 5.1A-B). This can then be represented as a binary code termed a *methylation tag*, where '1' represents a methylated CpG site and '0' an un-methylated CpG site (Figure 5.1C). Similarly, the methylation tags are also represented visually whereby a methylated CpG site is marked with a filled circle, and un-methylated CpG site by an un-filled circle (Figure 5.1D).

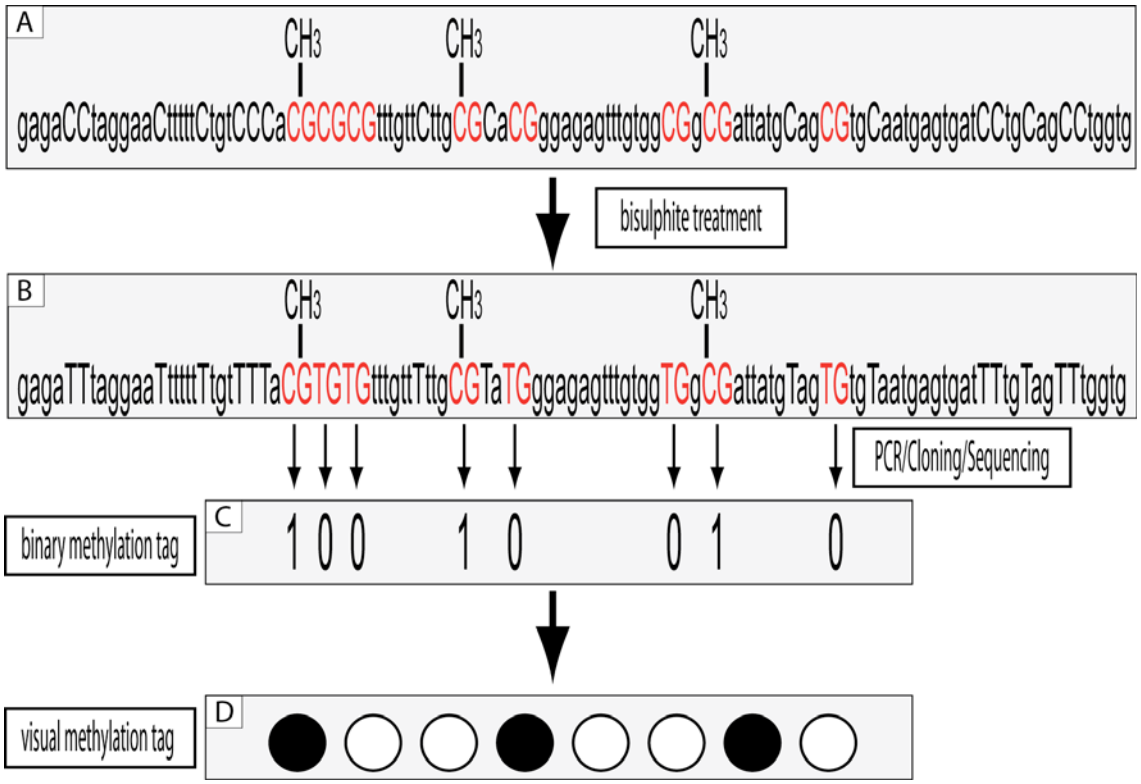

Figure 5.1 Bisulphite treatment allows the methylation status of gene-specific CpG islands to be sampled
(A): Sequence of the CpG island for the *CSX* gene which contains eight CpG sites that are either methylated (CH_3) or not. **(B):** Extracted DNA from a clone or other defined cell population is bisulphite treated: all un-methylated cytosine residues are converted to uracil. Subsequent PCR using primers specific for the methylated sequence and flanking the CpG site, followed by TA-cloning of individual PCR products and sequencing allows the methylation status to be sampled for that population. This can then be represented as a binary **(C)** or visual **(D)** methylation tag.

Since very small amounts of DNA were being bisulphite-treated when micro-dissecting clones from within crypts, sometimes perhaps only 400-500 cells, the ability of the assay to reliably measure methylation pattern diversity in small clonal populations was tested. Following DNA extraction of the clone or crypt, the DNA lysate was split into two aliquots and each was subjected to independent bisulphite-sequencing analysis. The two aliquots always produced near-identical methylation patterns (Figure 5.2A-D), confirming the sensitivity and reliability of the assay. Further, exhaustive cloning analysis showed that the summary statistics: the percent methylation and intra-crypt distance, were largely independent of the number of methylation tags analysed per crypt (Figure 5.3A-F), whereas the number of unique tags per crypt increased asymptotically with clone number (Figure 5.3G-I). These results are consistent with the validation of these techniques performed by previous authors [18], and demonstrated that the accuracy of the summary statistics are not significantly improved if more than 8 cloned PCR products are successfully sequenced.

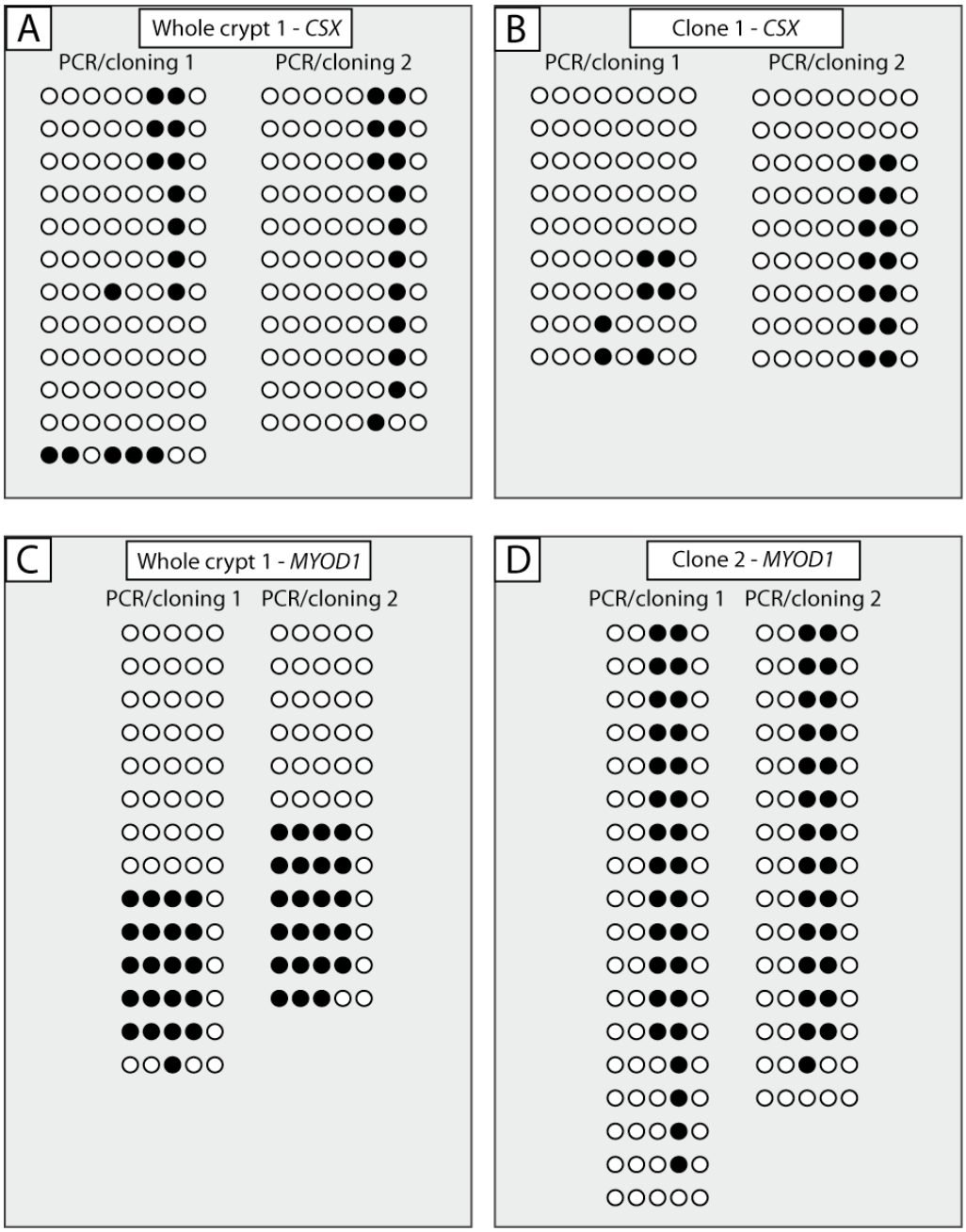

Figure 5.2 Sensitivity of clonal bisulphite sequencing assay
The ability of the assay to detect diversity within small clonal populations was determined by repeating the entire assay (bisulphite treatment, PCR and clonal sequencing) on individual crypt DNA lysates. The two repeats produced near-identical results, confirming the sensitivity of the assay. New methylation tags were revealed on the repeat, as was expected (see Figure 5.3). **(A) & (C):** analysis of whole crypt lysates for *CSX* and *MYOD1* respectively. **(B) & (D):** analysis of a partial crypt clone for *CSX* and *MYOD1* respectively.

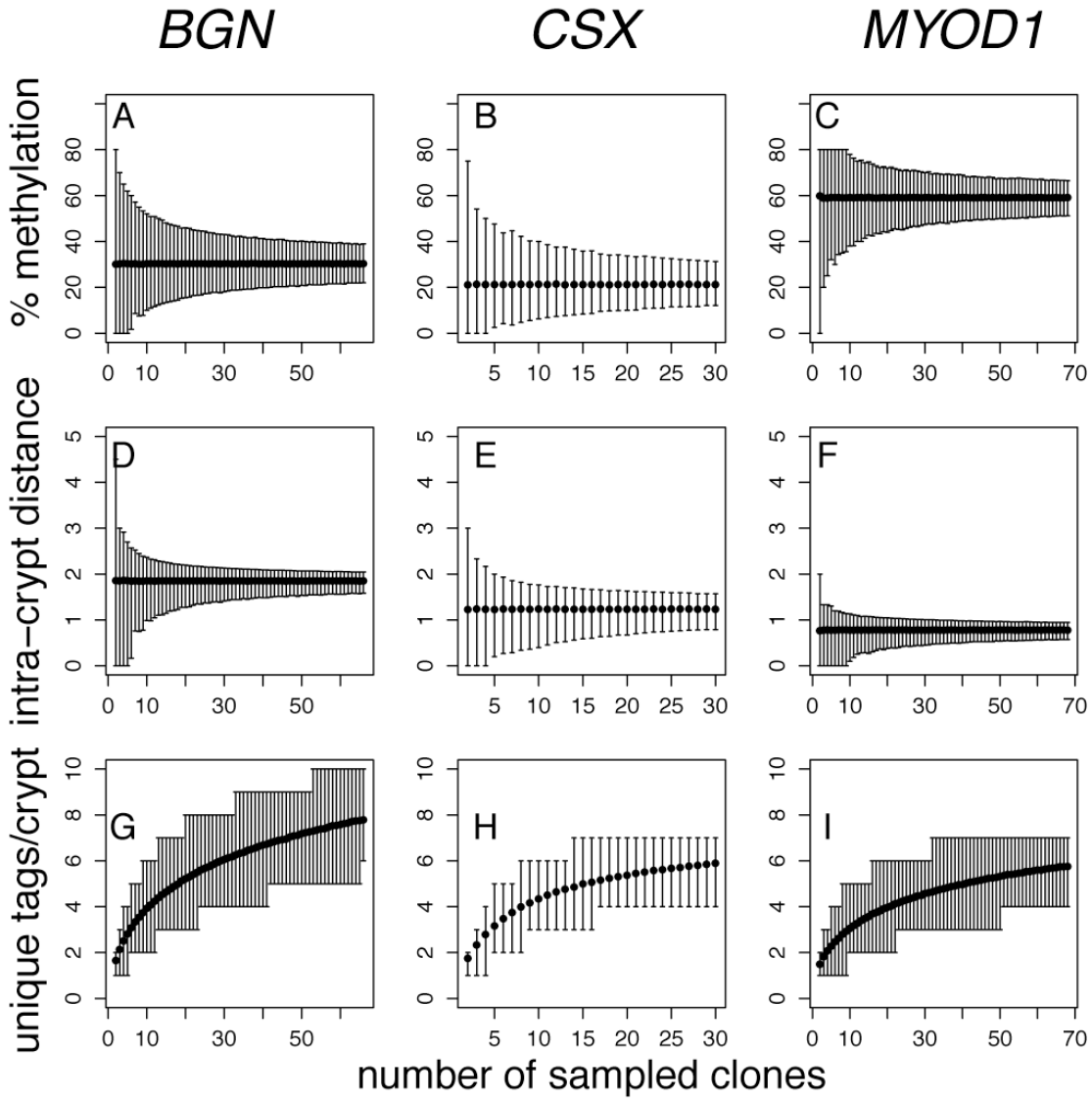

Figure 5.3 Cloning sufficiency
Exhaustive numbers of clones for each locus from a single crypt were sampled during the cloning phase of the bisulphite sequencing. A case-resampling bootstrap was applied to the sampled methylation tags to determine the reliability of the three summary statistics: percent methylation **(A-C)**, intra-crypt distance **(D-F)** and the number of unique tags per crypt **(G-I)**. The mean value of the percent methylation and intra-crypt distance were largely independent of the number of clones sampled, and the accuracy to which they could be determined was not significantly improved after 8 or so clones were sampled. Sampling large numbers of clones revealed new, rare, methylation tags, which may represent *de-novo* methylation errors in the non-stem cell compartment. Whiskers represent 95% confidence intervals.

5.3 Stem cell dynamics are unaltered by CCO-deficiency

In order to demonstrate that the stem cell dynamics of CCO-deficient crypts were no different to wild-type, CCO-proficient crypts, statistical analysis of crypt methylation tags comparing blue and brown crypts was performed. The percent methylation - a proxy for the mitotic age of the crypt stem cells, showed no significant difference between CCO-proficient and CCO-deficient crypts (Figure 5.4A-C), indicating that CCO-deficient crypts were derived from stem cells with the same mitotic age as wild-type crypts. Intra-crypt distance and the number of unique tags per crypt reflect stem cell number and tendency for stem cells to divide asymmetrically: no significant difference was seen between CCO-proficient and CCO-deficient crypts for either statistic (Figure 5.4D-I), confirming stem cell dynamics were unaltered by CCO-deficiency.

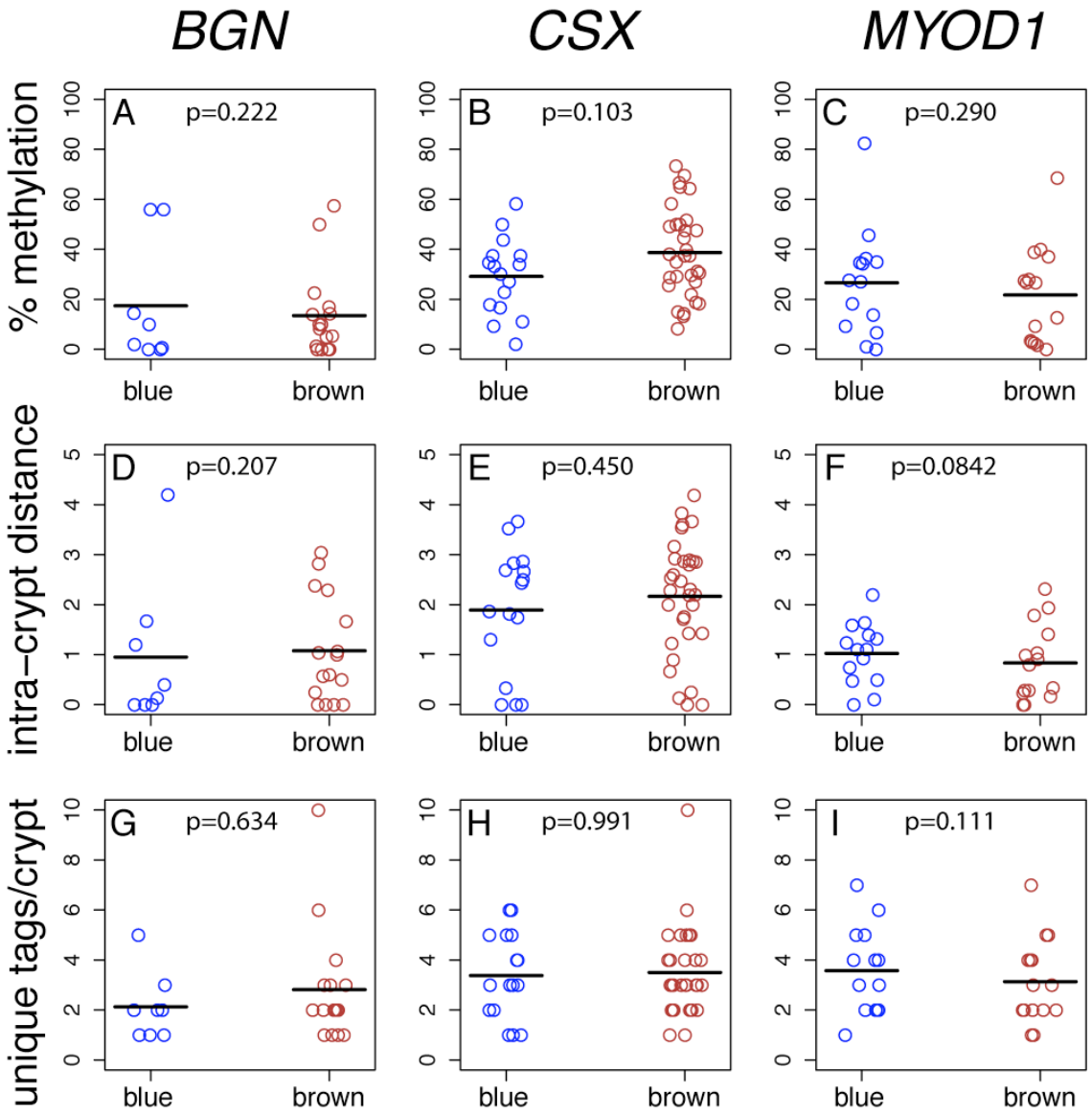

Figure 5.4 Epigenetic comparison of CCO-deficient and CCO-proficient crypts
The % methylation **(A-C)**, intra-crypt distance **(D-F)**, and number of unique tags per crypt **(G-I)** were compared between CCO-proficient (brown) and CCO-deficient (blue) crypts at each of the three CpG islands (*BGN*, *CSX* and *MYOD1*) analysed. CCO activity was detected by enzyme histochemistry. For all comparisons, there was no significant difference between the methylation patterns of CCO-proficient and CCO-deficient crypts indicating that stem cell dynamics were unaltered by loss of CCO expression. *BGN* n=25 crypts, *CSX* n=48 crypts, *MYOD1* n=21 crypts, mean 10 methylation patterns analysed per crypt.

5.4 Analysis of methylation patterns of colonic stem cell clones gives insights into niche dynamics

En-face serial sections of fresh frozen human colonic mucosa were stained for CCO activity and partially-deficient crypts were identified, genotyped and mapped as detailed previously (see Methods 2.4-2.5 and Chapter 3). The clonal, blue CCO-deficient cells and remaining brown, wild-type cell populations within the crypt were then separately laser capture micro-dissected (LCMD) and the methylation patterns at CpG islands of non-expressed genes of the two distinct cell populations sampled (see Methods 2.10). Clones of varying sizes within the crypt were sought in order to reflect different stages of clonal expansion and thus stem cell niche succession within the crypt. Statistical analysis was then performed (see Methods 2.11) to compare the methylation patterns of the clonal blue cells to the brown, wild-type population of the crypt, with clone size used as a measure of the time interval since formation of the clone: small clones were assumed to be newly formed.

Figure 5.5 illustrates a clone within a crypt where, after staining for CCO-activity, the mitochondrial genome of the blue cells was sequenced to confirm clonality (Figure 5.5A). The serial sections of the crypt were then photographed and a crypt map constructed (Figure 5.5B). Blue cells were then separately LCMD from the brown cells on each serial section (Figure 5.5C) and the methylation patterns at the CpG island of the non-expressed *CSX* gene sampled (Figure 5.5D). Mapping of the clones also allows the size of the clone within the crypt to be accurately measured by analysing the number of blue versus non-blue/black pixels in the crypt map. In this case the blue clone occupied 23% of the total crypt and had a relatively conserved methylation pattern at the *CSX* locus, with an intra-clone distance of only 0.36 compared to an intra-clone distance of 3 for the brown, wild-type cells within the crypt (Figure 5.5D), with only two distinct methylation tags detectable in the blue cells versus 4 in the brown cell population.

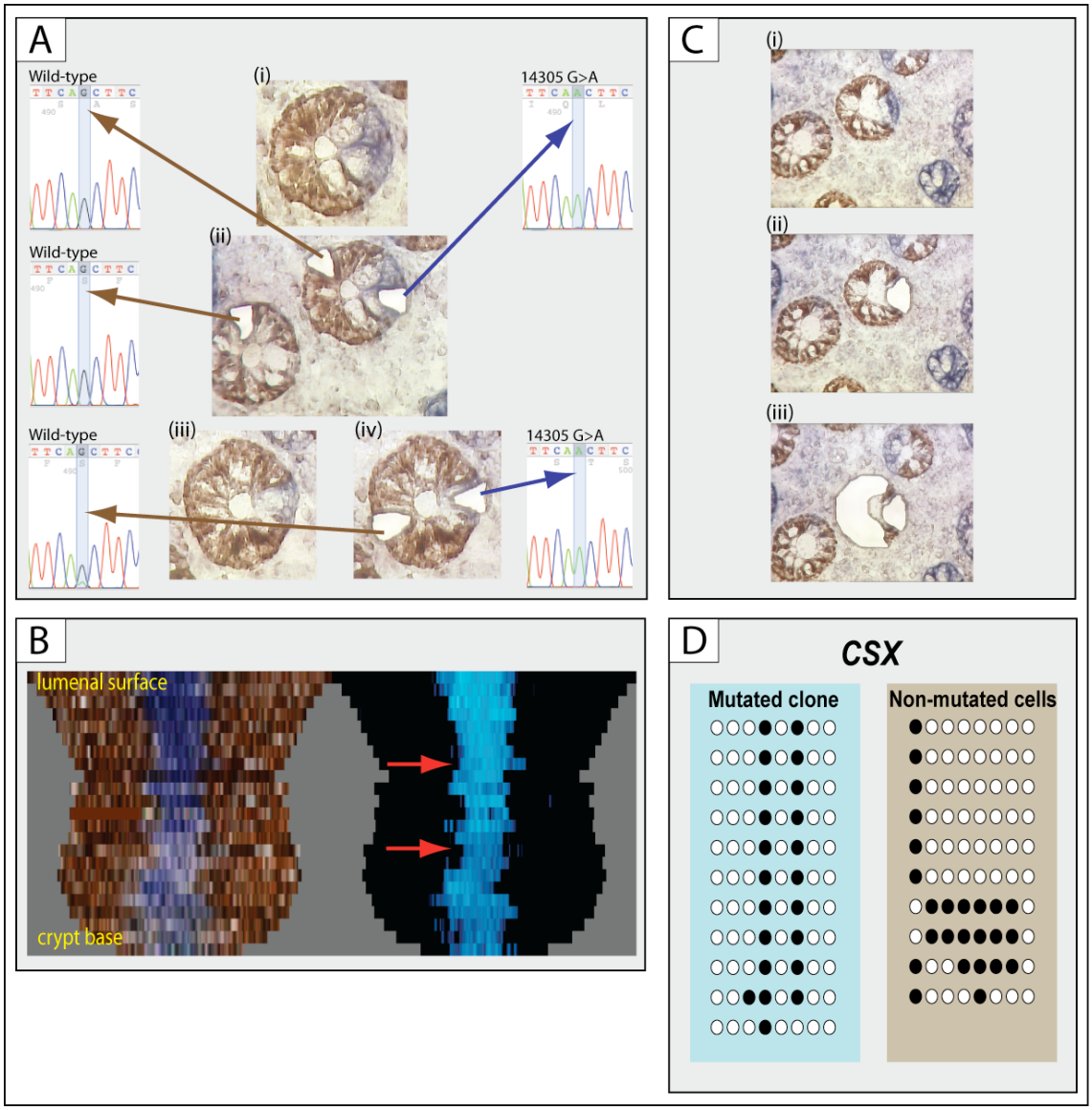

Figure 5.5 CCO-histochemistry enables dynamics of stem cell clones within crypts to be analysed

(A): CCO-histochemistry followed by LCMD and mtDNA sequencing enables a stem cell clone to be identified within a crypt. **(B):** Corresponding crypt map constructed using serial images of the crypt shown in **(A)**; red arrows indicate levels of the sections shown in **(A)**. The ratio of blue to black pixels indicated that the clone occupied 23% of the total crypt area. **(C):** The clonal blue cells were then separately LCMD from the brown cells within the crypt (ii & iii) in order to compare their respective methylation patterns. **(D):** The clone had a relatively conserved *CSX* methylation pattern (intra-clone distance 0.36) compared to the wild-type cells (intra-clone distance 3).

It was found that small CCO-deficient clones, which occupied less than one third of the crypt, had conserved methylation patterns, whereas the remaining wild-type population had diverse methylation patterns (Figure 5.6A-C) when analysing the intra-clone distance. However larger clones within crypts tended to have more diverse methylation patterns with intra-clone distances comparable to that of the wild-type population (Figure 5.6D): note that even when the brown cell population occupies only a small proportion of the crypt (< 30%), multiple distinct methylation tags are still picked up (Figure 5.7E-F), proving that the technique is sensitive enough to detect methylation tag diversity in very small cell populations.

Statistical analysis was then performed to look at the variance of methylation tags of clones as a function of the size of the clone within the crypt, in order to be able to infer the approximate time course of clonal expansion within the crypt, and hence the rate of niche succession. The *intra-clone distance* is a measure of the average similarity of all methylation tags within the CCO-deficient clone or wild-type cell populations of the crypt, and *number of unique tags* seen in a defined cell population is another measure of epigenetic diversity (see Methods 2.11). There was a general trend for more epigenetic diversity with larger clone size (Figure 5.7C-F) for *CSX* and *MYOD1*, significant for *CSX* ($p<0.003$) when looking at the most sensitive marker – the intra-clone distance; given the low rate of *de-novo* methylation, this increase in methylation pattern diversity with clone size indicates a low rate of niche succession in human colon crypts of many years. Again, importantly, the same trend was not seen when looking at the methylation tags of the brown, wild-type cells within the crypts (Figure 5.7C-F).

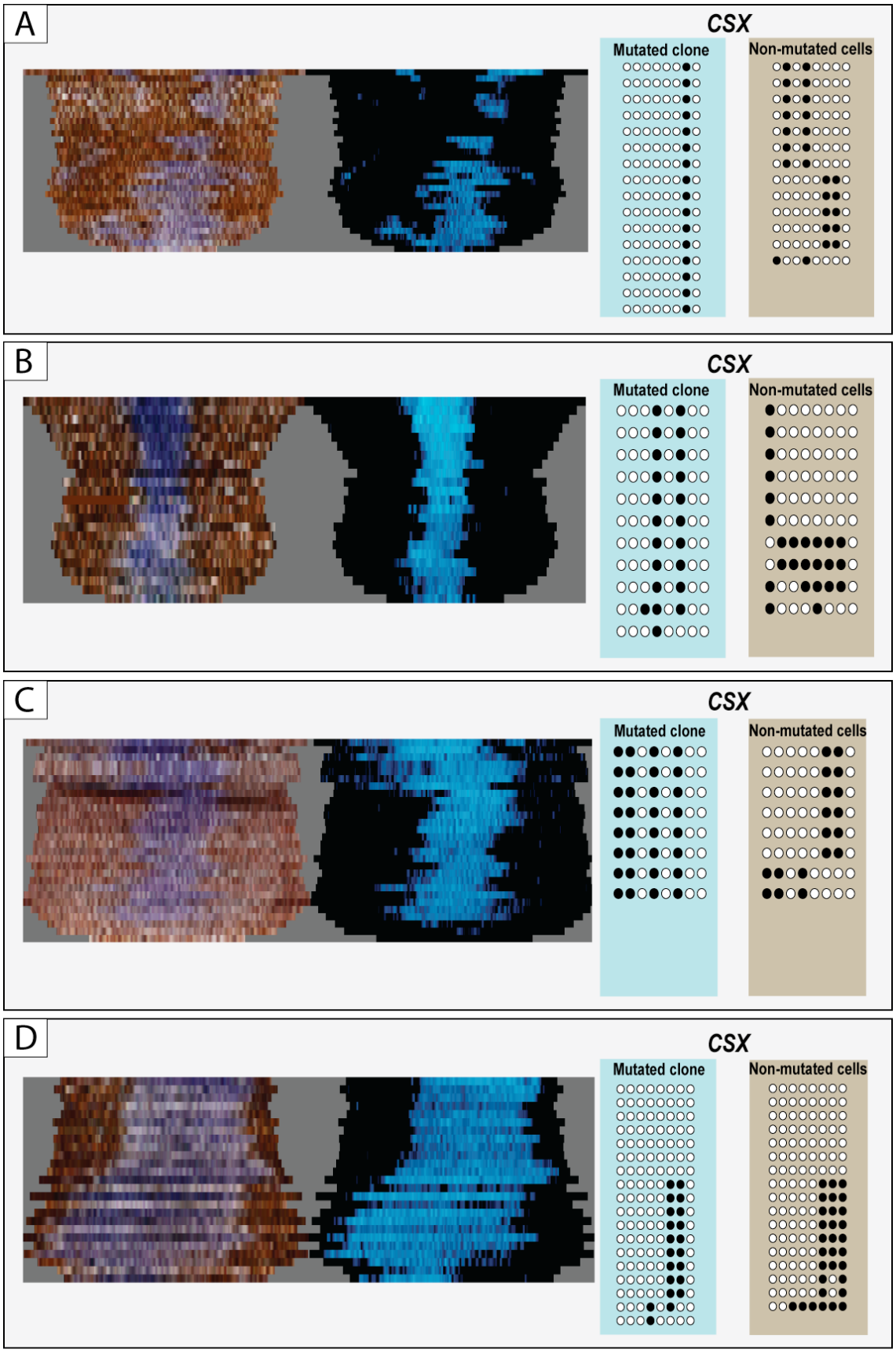
A
CSX
Mutated clone
Non-mutated cells
B
CSX
Mutated clone
Non-mutated cells
C
CSX
Mutated clone
Non-mutated cells
D
CSX
Mutated clone
Non-mutated cells

Figure 5.6 Methylation diversity increases with increasing clone size
For each panel the crypt map is shown on the left, with the methylation tags sequenced at the *CSX* locus for the respective blue and brown cell populations within the crypt on the right. Blue to black pixel ratios were used to calculate clone size. The clones in **(A) (B) & (C)** occupied 18%, 23% and 33% of the total crypt respectively and homogenous methylation patterns, whereas the wild-type population had diverse patterns (intra-clone distances 0 vs 2.3 for **(A)**, 0.36 vs 3 for **(B)** and 0 vs 2.1 for **(C)**). **(D)** is a larger clone totalling 54% of the crypt, and had developed diverse methylation patterns (intra-clone distance 1.43) comparable to that of the wild-type population (intra-crypt distance 1.91).

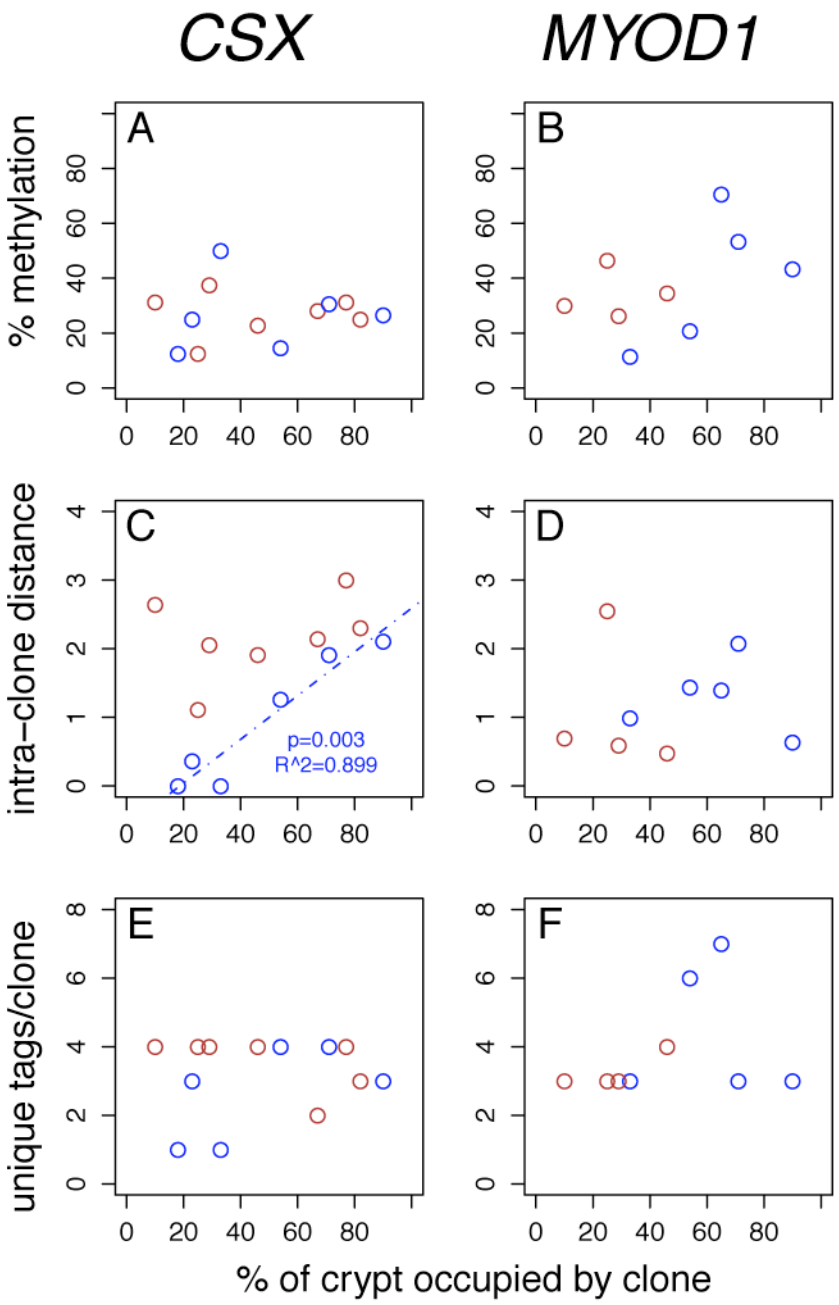

Figure 5.7 Statistical analysis of methylation tags of clones as a function of size within the crypt
(A-B): Percent methylation versus clone size. The percent methylation, a proxy for the mitotic age of the clone, was independent of clone size, indicating that the cell division rate was not increased in CCO-deficient clones. **(C-D):** Intra-clone distance versus clone size. Larger clones tended to have more diverse methylation patterns (*CSX*: p=0.003). **(E-F):** Unique tags/clone versus clone size. Data for *CSX* showed a non-significant trend towards larger clones containing a greater number of unique tags, but this was not clearly seen for *MYOD1*.

5.5 Dynamics of crypt fission in human colonic epithelium

It had been previously demonstrated that patches of adjacent CCO-deficient crypts are clonally derived from an ancestral CCO-deficient crypt that has undergone fission [77]. Thus, by then analysing the methylation tags of crypts that have undergone fission, it was hypothesised that the dynamics of this important mechanism of clonal expansion could be explored.

In Figure 5.8 LCMD and mtDNA sequencing of a patch of two adjacent CCO-deficient crypts demonstrates the presence of the same, clonal mtDNA point mutation in both crypts, confirming that they are clonally derived by fission (Figure 5.8A-B). However, these crypts have different, distinct methylation patterns (Figure 5.8C). Similarly in other patients, patches of adjacent CCO-deficient and therefore clonally-derived crypts had dissimilar methylation patterns, and similarity appeared independent of CCO-deficiency (Figure 5.8D-E). The *minimum distance* between the methylation patterns of two crypts, which compares each methylation tag in the first crypt to only its most similar methylation tag in the second crypt, indicates the epigenetic similarity between crypts. The minimum distance of adjacent CCO-deficient crypts were comparable to the distance between any pair of CCO-proficient crypts within the same colon for all loci studied (Figure 5.9D-F), indicating that adjacent CCO-deficient crypts appear as epigenetically unrelated as any other pair of random crypts sampled. The *inter-crypt distance*, which measures the average similarity of all methylation patterns in the first crypt to all the methylation patterns in the second crypt, did indicate that the *CSX* methylation patterns were more similar in adjacent CCO^- crypts ($p=0.001$) (Figure 5.9B). However, no significant difference was observed at the *MYOD1* locus ($p=0.8$) and the converse relation was seen for *BGN* tags ($p=0.007$) (Figure 5.9A & C). In addition, although the means for the inter-crypt distance at *CSX* were significantly different, there was a wide standard deviation.

Therefore, the clonal relation indicated by adjacent crypts both having CCO-deficiency was not mirrored by increased similarity of the methylation patterns of the crypts across all the loci analysed. As the rate of change in the methylation status at CpG sites is very low, thought to occur only every 1:200,000 cell divisions [18], the data suggests that crypt fission events occur infrequently in the normal colon and there are sufficient years between fission events for the methylation patterns to diverge.

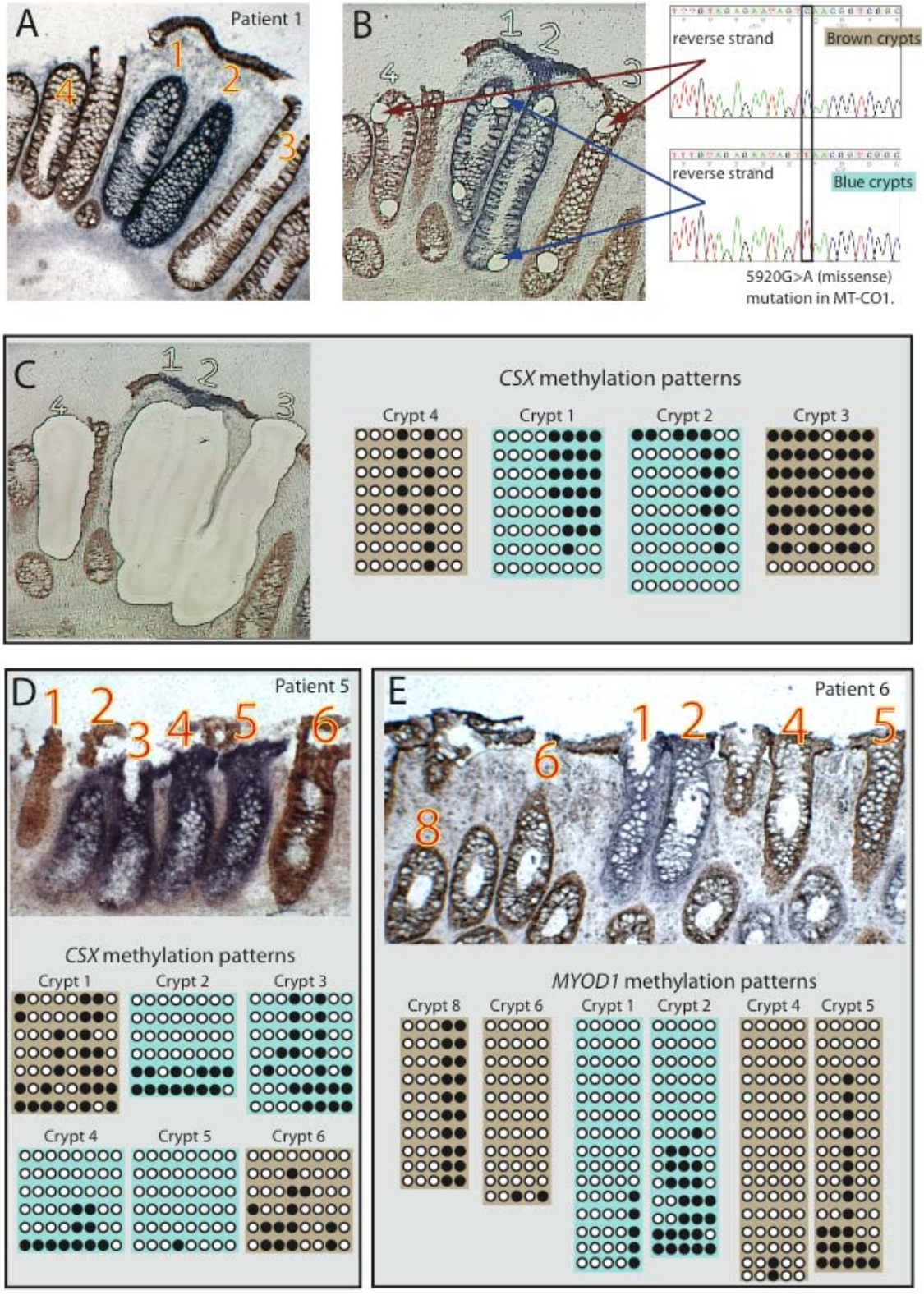
A
Patient 1
1
2
3
4
B
1
2
3
4
reverse strand
Brown crypts
reverse strand
Blue crypts
5920G>A (missense)
mutation in MT-CO1.
C
1
2
3
4
CSX methylation patterns
Crypt 4
Crypt 1
Crypt 2
Crypt 3
D
Patient 5
1
2
3
4
5
6
CSX methylation patterns
Crypt 1
Crypt 2
Crypt 3
Crypt 4
Crypt 5
Crypt 6
E
Patient 6
1
2
4
5
6
8
MYOD1 methylation patterns
Crypt 8
Crypt 6
Crypt 1
Crypt 2
Crypt 4
Crypt 5

Figure 5.8 Adjacent crypts have different methylation patterns irrespective of their clonal relation
(A): A section of morphologically-normal fresh frozen colon from patient 1 subjected to dual enzyme histochemistry for CCO and SDH. A patch of two CCO-deficient blue crypts (crypts 1 and 2) can be clearly seen. **(B)** Serial section showing dissected small cell areas. Cells from CCO-deficient crypts contained a homoplasmic 5920G>A missense mutation which was not present in the brown CCO-proficient crypts, confirming their common ancestry. **(C):** *CSX* methylation patterns from crypts 1-4: multiple methylation patterns are present in each crypt, as each crypt contains a number of long-lived stem cells that are methylated independently from one another; adjacent crypts have strikingly different methylation patterns - crypts 1 & 2 do not have the same methylation patterns despite being derived from a common ancestor crypt. **(D-E)**: Patch of four CCO-deficient crypts from patient 5 with *CSX* methylation patterns **(D)** and a patch of two CCO-deficient crypts from patient 6 with *MYOD1* methylation patterns **(E)**. All crypts within the patches have distinct methylation patterns despite being clonally derived by fission.

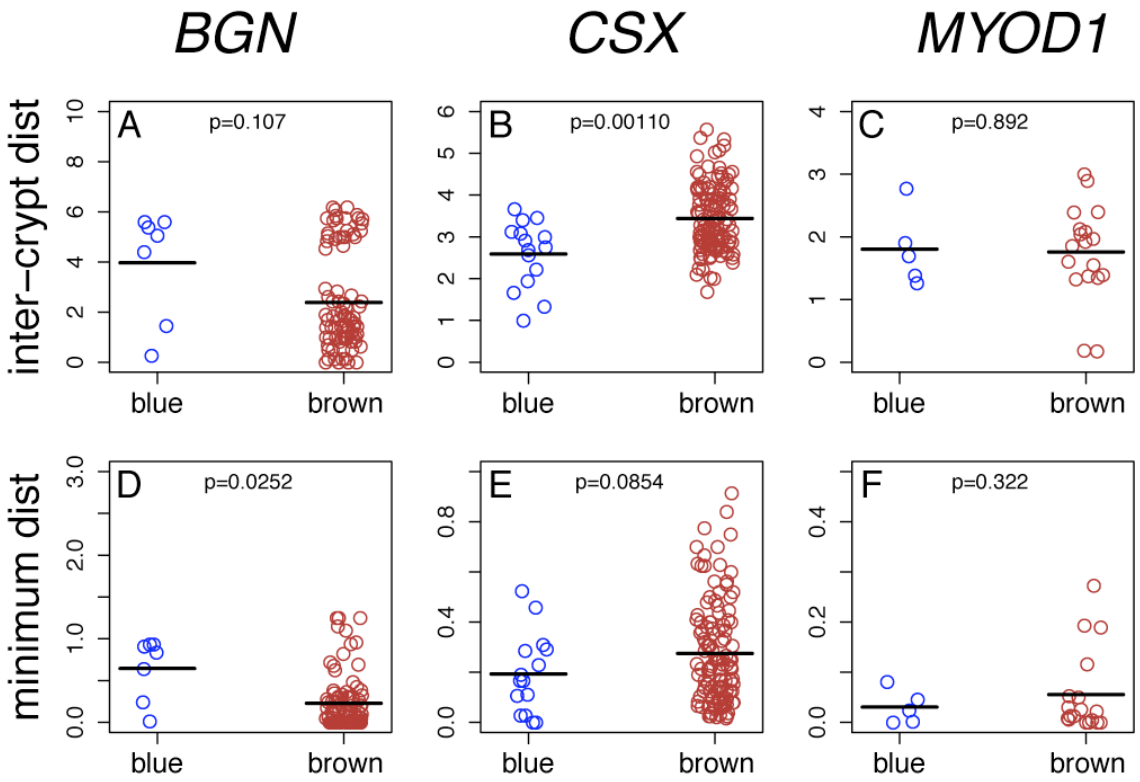

Figure 5.9 Methylation patterns of crypts that have undergone fission do not reflect crypt ancestry
(A-C): Comparison of the inter-crypt distance between two crypts in a patch (blue) and two random crypts not in a patch (brown) but within the same colon, for *BGN*, *CSX* and *MYOD1* respectively. Adjacent pairs of CCO-deficient crypts appeared more related by their *CSX* methylation patterns than any pairs of CCO-proficient crypts within the same colon (p=0.001), but the converse was true when *BGN* tags were examined (p=0.007) and no significant difference in the epigenetic relatedness was observed for *MYOD1* methylation patterns. **(D-E):** Minimum distance between two crypts within a patch and two random crypts not in a patch but within the same colon, for *BGN*, *CSX* and *MYOD1* respectively. All crypts appeared equally unrelated by this measure, regardless of whether the crypts were in a patch of CCO-deficient crypts.

5.5.1 Stem cell segregation during crypt fission causes methylation pattern divergence

Using *en-face* serial sections from patient 6, stained for CCO activity, partially-mutated bifurcating crypts were observed (Figure 5.10). Sections towards the crypt base revealed that one arm was CCO-deficient whilst the other was CCO-proficient, demonstrating that the segregation of stem cells at fission could result in the production of phenotypically distinct daughter crypts. Cells from the two arms where then separately micro-dissected and their methylation patterns compared. The CCO-proficient and CCO-deficient arms of each bifurcating crypt showed distinct patterns of methylation, indicating that segregation of stem cells during fission could result in daughter crypts with very different patterns of methylation (Figure 5.10A-B). The inter-crypt distances between the two arms of the bifurcating crypt were comparable to the inter-crypt distance between two unrelated crypts (A: *MYOD1* = 2.13; B: *CSX* = 3.95, *MYOD1* = 1.92). Notably, in Figure 5.10B, the *CSX* patterns of the two arms were very different, whereas the *MYOD1* patterns were similar, suggesting the stem cells in the ancestral crypt had homogeneous patterns of *MYOD1* methylation but heterogeneous *CSX* methylation.

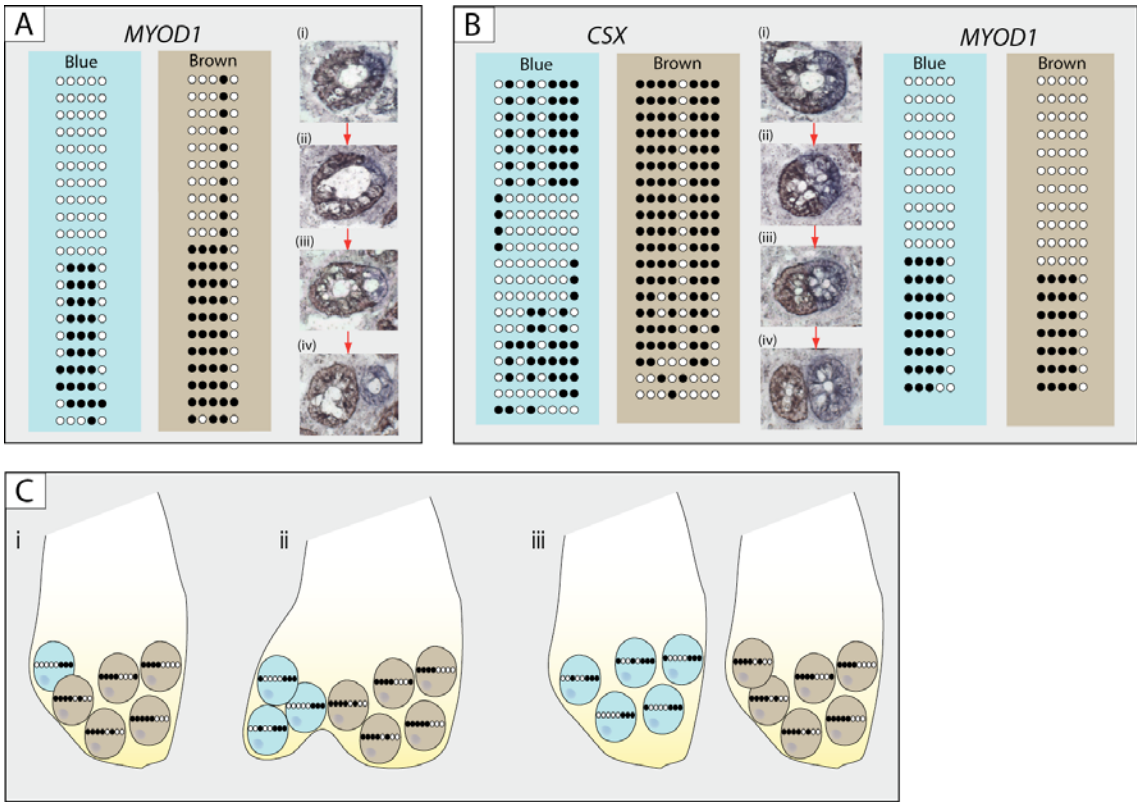

Figure 5.10 Segregation of stem cells at crypt fission causes rapid divergence of methylation patterns
(A-B): Serial images thorough two crypts in fission where one of the bifurcating arms contained blue, CCO-deficient cells, and the other brown, wild-type cells – a process termed *clonal purification* - with corresponding methylation patterns. In panels **(A)** and **(B)**, the top image is representative of the top third of the crypt: here the blue and brown cells were seen as a single crypt. In the middle third of the crypts, the blue cell population separated from the brown (middle two images) until in the lower third of the blue and brown cells were present in distinct arms of the crypt (bottom image). Methylation patterns of the blue, CCO-deficient arm vs the brown, wild-type arm are shown: In **(A)** the methylation pattern at *MYOD1* for the CCO-deficient arm was distinct from the wild type. In **(B)**, methylation patterns at *CSX* were again distinct between the CCO-deficient and wild-type populations, but almost identical at the *MYOD1* locus. **(C)**: Schematic representation of how asymmetric segregation of stem cells during crypt fission can result in rapid divergence of methylation patterns, even between the two bifurcating arms of a crypt in fission.

5.5.2 Mathematically modelling suggests there is at least 10 years between crypt fission events

A mathematical model of crypt stem cell methylation pattern dynamics was developed to assess how methylation patterns would diverge over time following fission, and the likely upper and lower time intervals between fission events that would result in patches of crypts clonally derived by fission appearing dissimilar. In their seminal paper, Yatabe et al [18] designed a mathematical model of crypt stem cell dynamics that was consistent with the methylation data at the *CSX*, *MYOD1* and *BGN* loci for normal individual human crypts. The Yatabe model was adapted in conjunction with Dr Trevor Graham (London Research Institute, Cancer Research UK) in order to assess crypt fission dynamics, further incorporating the potential effects of asymmetrical segregation of stem cells at fission.

The model was designed to investigate the rate of methylation pattern divergence in crypts dividing by fission. Crypts were assumed to contain exactly N synchronously dividing stem cells that tended to divide asymmetrically (probability p_1=0.95). Each stem cell had M CpG sites susceptible to methylation errors (methylation or de-methylation) at rate $\mu=2x10^{-5}$ per CpG site per cell division, as used by Yatabe et al [18] (Figure 5.11A). During fission at a predetermined time, stem cells were segregated between daughter crypts and symmetric divisions were assumed to restore the full stem cell compliment. Stem cells were either segregated *symmetrically*, whereby *N/2* of the total stem cells were segregated to each crypt, or *asymmetrically* whereby a single stem cell was inherited by one daughter crypt and the other *N-1* cells were placed in the second daughter crypt. Simulated daughter crypts with N=8 stem cells and M=8 CpG sites showed rapid divergence of methylation their patterns following symmetric fission of their parent crypt (Figure 5.11B).

In-silico power calculations were performed to indicate the ability of methylation to record ancestry. The inter-crypt distance between 10 pairs of related crypts was

compared to 10 pairs of unrelated crypts using a t-test. Statistical power was approximated repeatedly computing the model and counting the proportion of comparisons that gave a p-value less than 0.01. The power fell exponentially as a function of the time since crypt fission (Figure 5.11C). At 10 years post fission, with N=8, M=8 bi-allelic loci and symmetric crypt fission, a t-test would show pairs of related crypts to have smaller pair-wise inter-crypt distances than unrelated crypts at p<0.01 in only around 50% of tests. Asymmetric segregation of stem cells at fission had little effect on the ability to discern crypt relatedness. Divergence was slowed if an X-linked locus was probed (Figure 5.11C). A larger number of CpG sites (M=64) meant that daughter crypts were initially distinguishable by their methylation patterns, though after 10 years post fission, this record of ancestry was still rapidly lost (Figure 5.11C).

The mean inter-crypt distance between 8 crypts composing a patch found in a 90-year old colon (mean of inter-crypt distances computed between all possible pairs) that had formed through fission of ancestral crypts every 25 years was comparable to the mean inter crypt distance between 8 unrelated crypts (Figure 5.11D-E). Simulated power calculations suggested that a significant difference between the methylation pattern diversity of the eight related crypts to the eight unrelated crypts would be detected by a t-test in only 37% of comparisons. Thus a fission rate of once every 25 years would result in patches of clonally derived crypts being indistinguishable from random, neighbouring crypts by their methylation patterns.

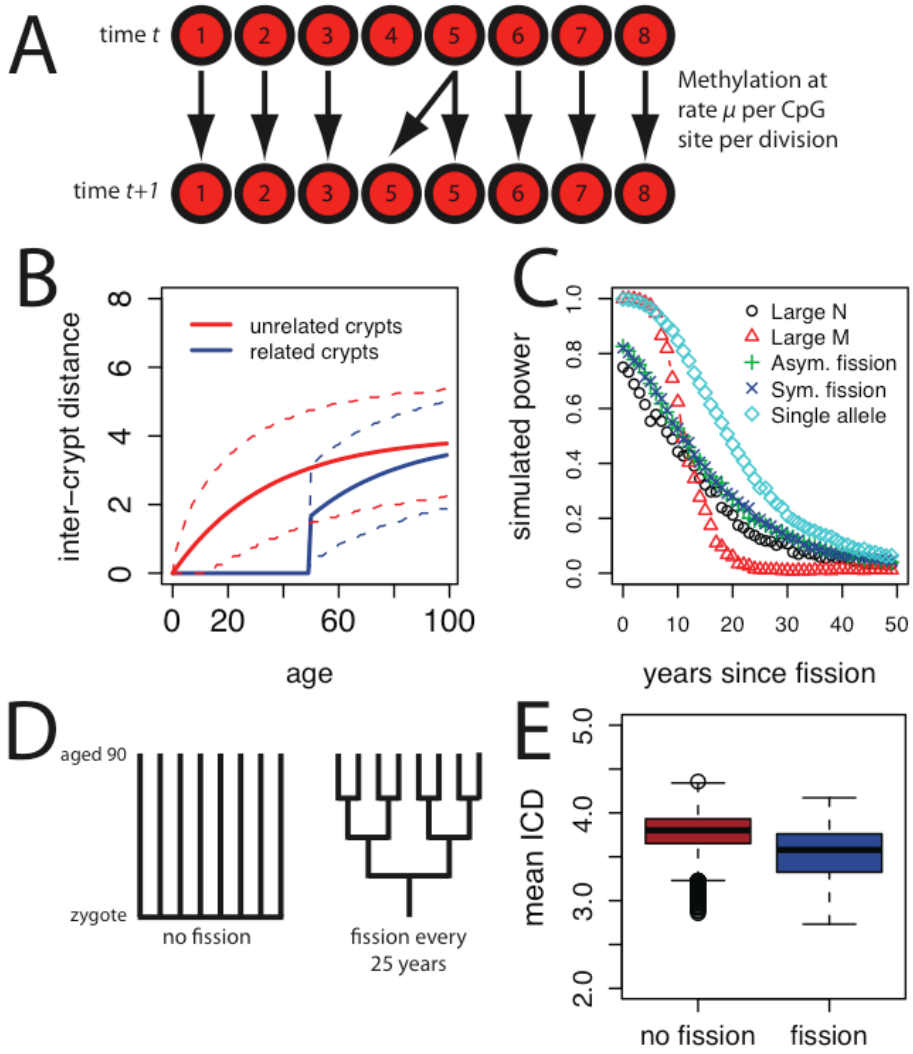

Figure 5.11 Crypt fission model of methylation pattern divergence
(A): Schematic of mathematical model. Crypts were assumed to contain a constant number of (N) stem cells. Occasional symmetric stem divisions producing two stem cell offspring were balanced by death of another stem cell lineage to maintain constant stem cell number. Methylation of the M CpG sites of each cell occurred at a rate of $\mu=2\times10^{-5}$ per site per division. **(B):** Inter-crypt distance between two unrelated crypts (red line), and two daughter crypts formed from a parent crypt that underwent fission (blue line) at $T_f=50$ years for $N=8$, $M=8$, symmetric crypt fission and a bi-allelic locus. Dashed lines represent the 2.5% and 97.5% quantiles of the simulated values. **(C)** Simulated power of t-test to show that 10 pairs of daughter crypts (derived from an ancestral crypt that has divided) have a smaller inter-crypt distance than 10 pairs of unrelated crypts. Discriminatory power of the test falls rapidly with the time since crypt fission. Large M or N further reduce the ability of the test to discriminate at 10 years post fission. **(D-E):** The methylation patterns of 8 unrelated crypts were compared to the methylation patterns of 8 crypts that formed through repeated rounds of fission (every 25 years) of an ancestral crypt. Related crypts in a patch showed similar dissimilarity of their methylation patterns to the unrelated crypts.

Experimental measurements suggest that the methylation error rate of non-expressed loci may be as high as μ=0.005, although precise estimates of the methylation error rate are locus dependent [180, 181]. Using this higher methylation error rate, the simulations showed extremely rapid divergence of daughter crypts following fission, and concomitant significant reductions in the power of a t-test to distinguish the methylation pattern diversity of 8 pairs of related and unrelated crypts.

5.6 Discussion

Individual human CpG sites are estimated to change their methylation status at a low rate (approximately $2x10^{-5}$ per CpG site per cell division) [18]. However, CCO-deficient clones occupying large areas of a crypt (>50%) showed multiple, distinct methylation patterns (Figure 5.6 and Figure 5.7), indicating that niche succession occurs sufficiently slowly to permit *de-novo* methylation, likely over many years. Direct inferences of the niche succession rate from methylation patterns requires a more precise knowledge of both the methylation rate and stem cell division rates. Previous human studies have supported the concept that niche succession occurs over many years: the proportion of wholly O-acetyltransferase (OAT) defective crypts in radiotherapy treated $OAT^{-/+}$ individuals increases slowly over 30 years post radiation therapy, indicating OAT^- cells can take over a decade to reach dominance in some crypts [74].

Adjacent CCO-deficient crypts tend to be derived from a single founder crypt that undergoes fission in the adult colon [77]. Despite this recent relationship, adjacent CCO-deficient crypts usually appeared no more related by their methylation patterns than physically disparate crypts (Figure 5.8 and 5.9). Thus the time periods between crypt fission events are sufficiently long for methylation patterns of two daughter crypts to diverge. This is consistent with the study by Kim et al [168] where randomly selected adjacent crypts, in the absence of a clonal marker, appeared no more similar by their methylation patterns than crypts up to 15cm away.

How long is the methylation pattern divergence time of related crypts? An upper bound can be estimated by considering the time elapsed since a patch of related crypts was formed. The patches analysed in this chapter were from patients aged between 68-78 years at the time of resection. CCO-deficient crypts are rarely observed in colonic mucosa from individuals aged less than 40, therefore we can assume that a patch of two CCO-deficient crypts formed as a result of a fission event in the bowel aged more than 40 years [77]. This implies that the crypts within a patch are separated by at most 35 years from their most recent common ancestor. However, the mathematical analysis predicts a significantly lower bound for the divergence time: even with a low methylation error rate, after approximately 10 years the probability of distinguishing ten pairs of related crypts from 10 pairs of unrelated crypts has dropped to 0.5 (Figure 5.11), though this prediction can be considered no more than an estimate due to the uncertainty about the precise methylation error rate. This rapid divergence of methylation patterns is the result of stem cell dynamics within the niche and can be exaggerated by stem cell segregation at fission. Modelling the methylation patterns of a parent crypt in the zygote that has divided by fission every 25 years, demonstrates that a patch of 8 crypts in a 90 year old would appear no more related by their methylation patterns than a patch of 8 random, unrelated crypts (Figure 5.11) - consistent with the biological data. Thus, from the data collected here it appears that crypt fission in the normal human colon occurs with a lower interval of approximately 10 years and an upper interval in the region of 25 years.

Choice of methylation-susceptible locus also influences ability to infer clonality. Notably, in this study, *CSX* tags appeared to record ancestry more effectively than *MYOD1* tags: Adjacent CCO-deficient crypts had more similar *CSX* methylation patterns than unrelated crypts, and *CSX* methylation pattern diversity scaled with clone size in partially-mutated crypts, whereas these trends were not observed for *MYOD1*. This differential ability to record ancestry is likely to result from both the locus and CpG-site specific (de-)methylation rates. Perhaps counter-intuitively, simply increasing the number of CpG sites probed does not, in the long-term, lead to a decrease in the average pair-wise distance between the methylation tags from related crypts, as

epigenetic drift through random methylation is more marked when a greater number of CpG sites are considered. Summary statistics to describe methylation pattern diversity must also be chosen with care. For example, sampling more tags from a clone will find occasional new methylation patterns (Figure 5.3), which may represent de-novo methylation events in the non-stem cell compartment. Average measures, such as the intra-clone distance, are more robust descriptors of methylation pattern diversity.

A rapid change in the phenotype of crypts due to the segregation of cells at fission is a process termed *clonal purification*. Examples of clonal purification have been previously observed in the budding crypts of mice with induced somatic mutations of the *Dlb-1* locus [17] and in mutagen treated mice heterozygous for a G6PD mutation [71]. The epigenetic diversity (as measured by the inter-crypt distance) between the two arms of crypts in fission was comparable to that of two unrelated crypts (Figure 5.10), this could be due to asymmetric stem cell segregation or increased stem cell number or cycling within crypts in fission. Kim and Shibata [168] observed that, on average, methylation patterns of crypts in fission appeared as diverse as the methylation of two unrelated crypts, concluding that fission may often disrupt the epigenetic record in normal crypts. It has been shown that increased cellular proliferation is not associated with increased crypt fission events [83, 182], thus the increased diversity between the bifurcating arms could be evidence that expansion of the stem cell zone initiates fission events in the colon, as has been proposed to explain the increased crypt fission rates observed in *APC* mutated crypts [83, 85]. It is unclear whether the segregation of the CCO-deficient portion of the crypt is a random process, these were very rare events and no partially-deficient crypts in fission were observed where there was mixed populations of cells in the bifurcating arms.

The dynamic nature of methylation patterns severely restricts their use as a clonal marker *per-se*. However, here clonality has been inferred by the presence of clonal point mutations in mtDNA, a robust marker, and the additional analysis of methylation patterns of non-expressed genes provide an attractive means to estimate the rates of

recent clonal expansions in human tissues. Precise estimates are precluded, however, due to the inherent uncertainty over exact cell division and methylation error rates.

Chapter 6. Dynamics of human colorectal sporadic adenoma initiation and growth

6.1 Introduction and aims

Most of our current data on the dynamics of adenoma growth is based on longitudinal endoscopic and barium studies (see Introduction 1.5). These did not look at the detailed genetics of the adenomas they were assessing, and observed that adenomas less than 1 cm appear to be static in their growth, with many demonstrating a tendency to regress.

The aims of this project were to use the techniques developed in the previous chapters for assessing clonal expansion in the normal epithelium to investigate, at a microscopic and molecular level, the dynamics of adenoma initiation, growth and clonal expansion. Crypt fission is thought to be the predominant mechanism by which a mutated crypt can clonally expand within the epithelium to form an adenoma [93], and it has been shown that crypt fission rates are elevated in hyperplastic polyps and adenomas [94]. If the mechanism by which a single mutated stem cell expands within a crypt is niche succession, with subsequent monoclonal conversion and crypt fission, it was hypothesised that these events would be significantly up-regulated within adenomas to enable accelerated clonal expansion.

Previous studies have established that adenoma growth is accompanied by a step-wise accumulation of oncogenic mutations over time [113, 114], and that genetically distinct sub-clones can be found within adenomas [106, 107]. It was thus hypothesised by using both mutations in mtDNA and nuclear genes as clonal markers, combined with methylation analysis of CpG islands within non-expressed genes, the dynamics of clonal evolution within adenomas could be investigated.

6.2 Mutation screening and global DNA methylation status

Fresh frozen sporadic human adenomas and polyps were collected as described (See Methods 2.1). Only early adenomas with low-grade dysplasia were selected, along with hyperplastic polyps, as the crypt structures are relatively maintained and there are less confounding factors, such as CIMP, global hypo-methylation or chromosomal instability that are found in more advanced adenomas and cancers [114, 123, 135, 183]. The adenomas and polyps were assessed for size and histology, and screened for mutations in *APC, KRAS, BRAF, TP53* and loss of heterozygosity (LOH) in 5q (*APC*) and 17p *(TP53*); MSI status was also determined (see Methods 2.12). This was important in order that the dynamics could be related to the specific genomic changes and also to try to identify those adenomas where it may be possible to define sub-clones with differing oncogenic mutations, specifically in *KRAS* or *APC*, as has been shown in previous studies [106, 107]. When mutations were detected on screening, this was confirmed on each individual crypt before determining the methylation status of the CpG islands in *CSX* and *MYOD1*. Due to the conflicting results obtained with *BGN* obtained when looking at normal tissue this locus was not used for further analysis.

The commonly mutated areas within the *APC, KRAS, BRAF* and *TP53* genes were sequenced by PCR (see Methods 2.12.1-2.12.4), and MSI status determined using a multiplex PCR for the *BAT25* and *BAT26* mononucleotide repeats (see Methods 2.12.5). The results are summarised in Table 6.1: 8/9 (88%) adenomas harboured mutations within the MCR of the *APC* gene, and 2/9 (22%) had additional *KRAS* mutations. A possible confounding factor when comparing the methylation status at CpG loci between lesions is that increased methylation may be due to global increased methylation rates within the adenoma, rather than a difference in dynamics. The CpG island methylator phenotype (CIMP) (see Introduction 1.4), a mechanism whereby progression of the adenoma is associated with methylation of the promoter regions of a number of genes, is closely associated with BRAF mutations and microsatellite instability [135, 146, 147] (MSI). Therefore all lesions were screened for these, and found to be BRAF wild-type and microsatellite stable (MSS) (Table 6.1 and Figure 6.1), thereby

making it less likely elevated methylation rates within these adenomas would be confounding the results.

Lesion	Location, Size & Histology	Molecular screening				
		APC	***KRAS***	***BRAF***	***TP53***	**MSI**
Polyp 1 **Patient 10**	Sigmoid colon 1.1cm;hyperplastic	WT	WT	WT	WT	MSS
Adenoma 1 **Patient 11**	Ascending colon 2.2cm;TVA;LGD	c.4348 C>T p.R1450*	c.35 G>A p.G12D	WT	WT	MSS
Adenoma 2 **Patient 12**	Sigmoid colon 2.2cm;TA;LGD	c.3964 G>T p.E1322*	WT	WT	WT	MSS
Adenoma 3 **Patient 13**	Rectum 1.3cm;TVA;LGD	c.3956delC p.A1299fs*1	c.38 G>A p.G13D	WT	WT	MSS
Adenoma 4 **Patient 14**	Sigmoid colon 1.2cm;TVA;LGD	c.4118_4126delCTGAACC p.P1371fs*10	WT	WT	WT	MSS
Adenoma 5 **Patient 15**	Splenic flexure 2cm;TVA;LGD	c.4474_4475delGCinsT p.A1492fs*14	WT	WT	WT	MSS
Adenoma 6 **Patient 16**	Sigmoid colon 1.7cm; TVA;LGD	c.3925_3929delGAAAA p.E1309fs*4	c.35 G>A p.G12D	WT	WT	*NR*
Adenoma 7 **Patient 17**	Sigmoid colon 1.4cm;TVA;LGD	WT	WT	WT	WT	MSS
Adenoma 8 **Patient 18**	Sigmoid colon 1.1cm;TA;LGD	c.4461_4468delTTTATTAC p.T1487fs*6	WT	WT	WT	*NR*
Adenoma 9 **Patient 19**	Rectum 1.2cm;TA;LGD	c.3927_3931delAAAGA p.E1309fs*3	WT	WT	WT	*NR*

Table 6.1 Summary of mutation screening results
All lesions were screened for tumorigenic changes known to commonly occur in human adenomas and polyps. Initial screening was done from scrapes of 3 serial sections through the lesion (see Methods 2.2), when a mutation was found this was confirmed by sequencing multiple individual crypts from that lesion. Abbreviations: TVA tubulo-villous adenoma; TA tubular adenoma; LGD low-grade dysplasia. *APC* Adenomatous Polyposis Coli gene; *KRAS* v-Ki-ras2 Kirsten rat sarcoma viral oncogene; *TP53* Tumour protein 53 gene; MSI microsatellite instability; WT wild-type; MSS microsatellite stable; *NR* no result.

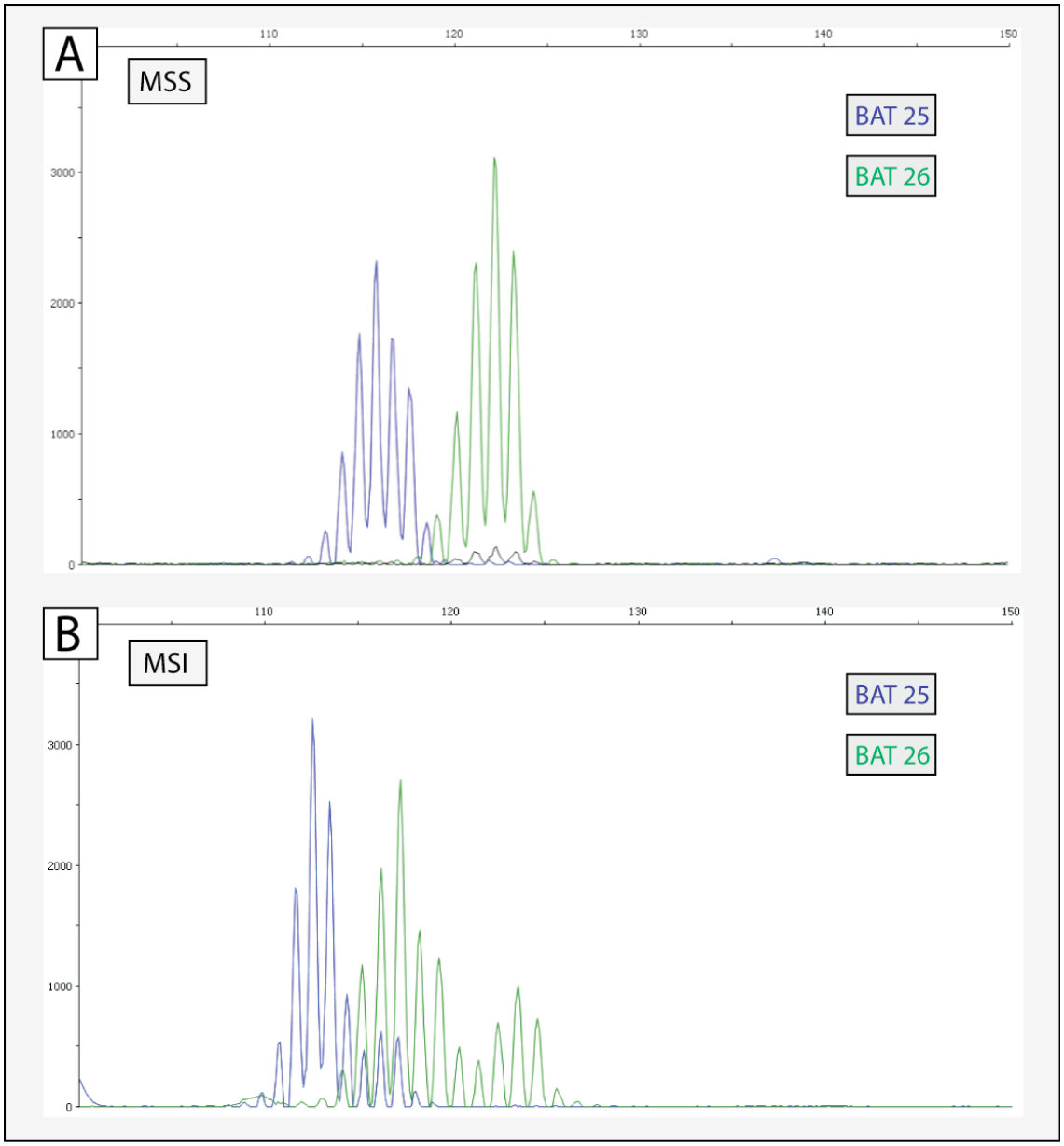

Figure 6.1 MSI status analysis
Multiplexed PCR for BAT25 and BAT 26 repeats. **(A):** All samples were MSS (example trace given from adenoma 1). **(B):** Example of an MSI trace from a Crohn's colitis associated cancer. Abbreviations: MSI microsatellite instability; MSS microsatellite stable.

All lesions were screened for LOH at 5q (*APC*) and 17p (*TP53*) using a panel of microsatellite markers; the data is summarised in Table 6.2 and examples of the LOH analysis demonstrated in Figure 6.2. Mutations in *TP53* and 17p LOH tend to occur later on in adenoma development, and are associated with the progression to high-grade dysplasia and cancer [113, 114, 144]. All the adenomas in this study were of low-grade dysplasia, thus, unsurprisingly no mutations in *TP53* or 17p LOH were detected (Table 6.1 and Table 6.2). Once *APC* is mutated, LOH of the wild-type allele may the process by which function of the second allele is lost [184]. None of the adenomas demonstrated LOH at *APC* by the markers used here (Table 6.2). It may be that the second hit in *APC* was outside of the MCR, or missed during the analysis, or that the area of LOH was not detected by the microsatellite markers.

Lesion	**LOH marker**			
	APC (5q/D5S346)	*APC* (5q/D532001)	*APC* (5q/D5S489)	*TP53* (17p/D17S1832)
Polyp	No LOH	*NI*	No LOH	*NI*
Adenoma 1	No LOH	No LOH	No LOH	No LOH
Adenoma 2	No LOH	No LOH	No LOH	*NI*
Adenoma 3	No LOH	No LOH	No LOH	*NI*
Adenoma 4	*NI*	No LOH	*NI*	*NI*
Adenoma 5	*NR*	*NR*	*NR*	*NR*
Adenoma 6	*NR*	*NR*	*NR*	*NR*
Adenoma 7	No LOH	*NR*	*NR*	*NR*
Adenoma 8	No LOH	No LOH	No LOH	*NI*
Adenoma 9	No LOH	No LOH	No LOH	No LOH

Table 6.2 Summary of LOH results
Multiplexed PCR was performed for a panel of microsatellite markers located within or near the *APC*, and *TP53* genes. LOH was considered present if the area under one allelic peak was more than twice that of the other, after normalising the peak areas relative to the constitutional DNA. No LOH was detected where markers were informative Abbreviations: LOH loss of heterozygosity; *NI* not informative (constitutionally homozygous marker); *NR* no result (no normal control available or failed PCR).

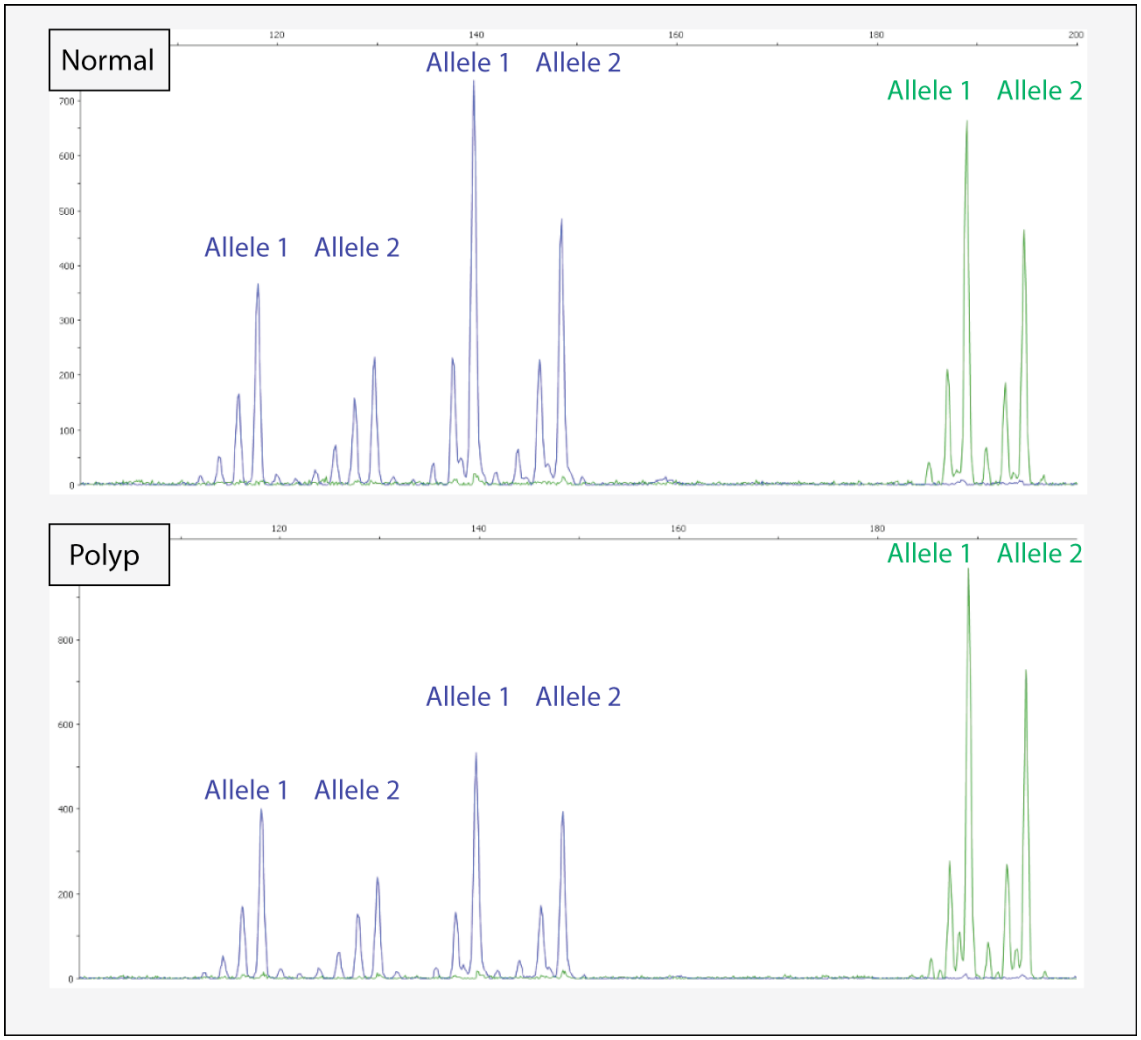

Figure 6.2 Multiplex PCR LOH analysis
The ratio of the 2 allelic peaks is calculated for the adenoma or polyp and then normalised against that for the normal tissue; a normalised ration of <0.5 or >2.0 indicates a two-fold change and is taken as indicating the presence of LOH. GeneScan® traces for normal and polyp tissue from polyp 1 (patient 10) are shown, demonstrating absence of LOH at three microsatellite markers.

In order to further assess the lesions for global hyper-methylation that might potentially confound the analysis of CpG islands in adenomas, the methylation status at CpG islands of long interspersed nucleotide elements (LINE-1) within genomic DNA, shown to be a surrogate marker of global genomic DNA methylation [135, 149, 172], was quantified in order to see if there was any evidence for genome-wide hyper-methylation within the adenomas compared to normal controls (see Methods 2.13). The results are summarised in Figure 6.3, importantly there was no significant difference between the mean methylation levels across the four LINE-1 CpG sites sampled when comparing normal versus adenoma tissue (p=0.591), or between paired adenoma and normal colonic mucosal tissue from the same patient (p=0.5469).

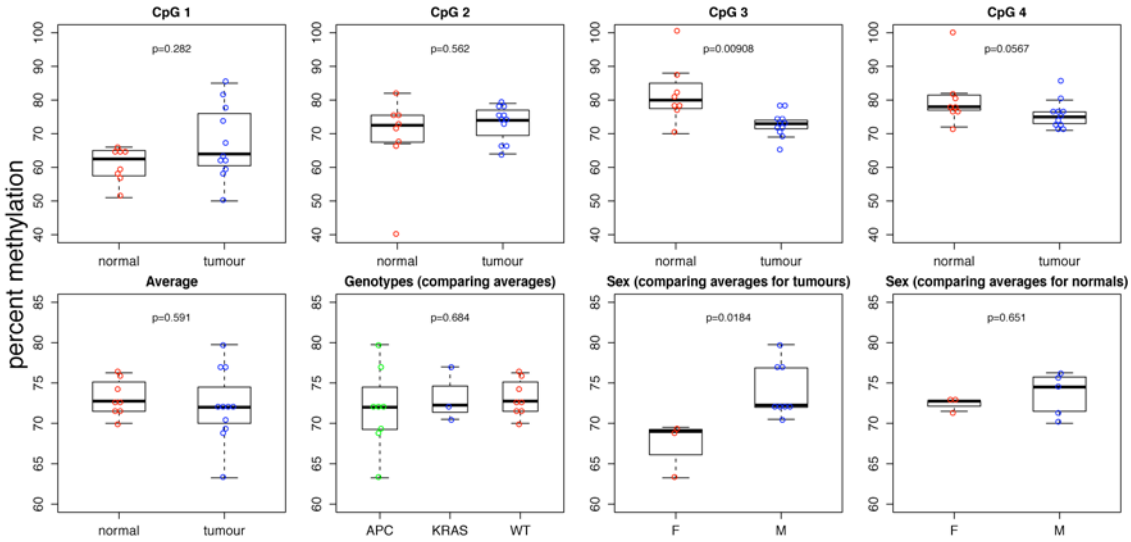

Figure 6.3 CpG methylation status of LINE-1 elements as measured by Pyrosequencing®
The methylation level at each of the four CpG sites was quantified by pyrosequencing® (see Methods 2.13). There was no significant difference in methylation between normal and adenoma tissue, or between adenomas with an *APC* versus *APC* and *KRAS* mutation. Significantly lower methylation at LINE-1 CpG sites for adenomas from female patients was observed versus adenomas from male patients, a trend that was not seen in normal tissue. Statistical analysis done using a Mann-Whitney U-test for all comparisons, other than genotypes where a Kruskal-Wallis test was used; data plotted as a boxplot. There was no significant difference (p=0.5469) in methylation at the CpG sites between paired normal and adenoma tissue from the same patients, when analysed using Mann-Whitney U-test.

The adenoma tissue for the LINE-1 analysis was macro-dissected from slides containing only adenomatous crypts, and no normal epithelial tissue. In order to check whether stromal tissue within the adenoma might be confounding the results, for adenoma 2 both a macro-dissected and LCMD sample of adenomatous crypts only were run in parallel; the mean methylation level across the four CpG sites was 72.08% for the macro-dissected tissue and 72.23% for the LCMD sample, suggesting that any normal stromal cells present within the macro-dissected adenoma samples did not significantly affect the Pyrosequencing® analysis.

6.3 Large clonal patches that remain similar by their methylation patterns can de identified within human adenomas and polyps

After staining for CCO-activity, multiple, large patches of CCO-deficient crypts were seen within some of the lesions. Laser capture micro-dissection and PCR of the mitochondrial genome (see Methods 2.4) was then performed on the blue and surrounding brown crypts in order to confirm clonal somatic point mutations in mtDNA. Using DNA from the same crypt, methylation patterns at the CpG islands of *CSX* and *MYOD1* were determined using clonal bisulphite sequencing and statistical analysis, performed as before, in order to infer the dynamics of these clonal expansions (see Methods 2.10 and 2.11).

The polyp from patient 10 was a hyperplastic polyp of the goblet cell variety with some serrated features. There were 3 distinct CCO-deficient patches that all had differing clonal mutations in their mtDNA (Figure 6.4A-E). The whole mitochondrial genome was sequenced for blue crypts in each patch in order to confirm that they were indeed clonally distinct from each other by their mtDNA; mtDNA of cells from adjacent brown, wild-type crypts to each patch was also sequenced. Crypts within adenomas and polyps are atypical and often convoluted, therefore each crypt that was LCMD within a

patch was carefully followed on serial sections throughout the lesion to ensure that is was indeed distinct, and that the same crypt was not simply being micro-dissected on multiple occasions. One of the patches harboured four clonal mutations (Figure 6.4C), and the patches all contained a large number of clonally related crypts, with at least 6 crypts per patch.

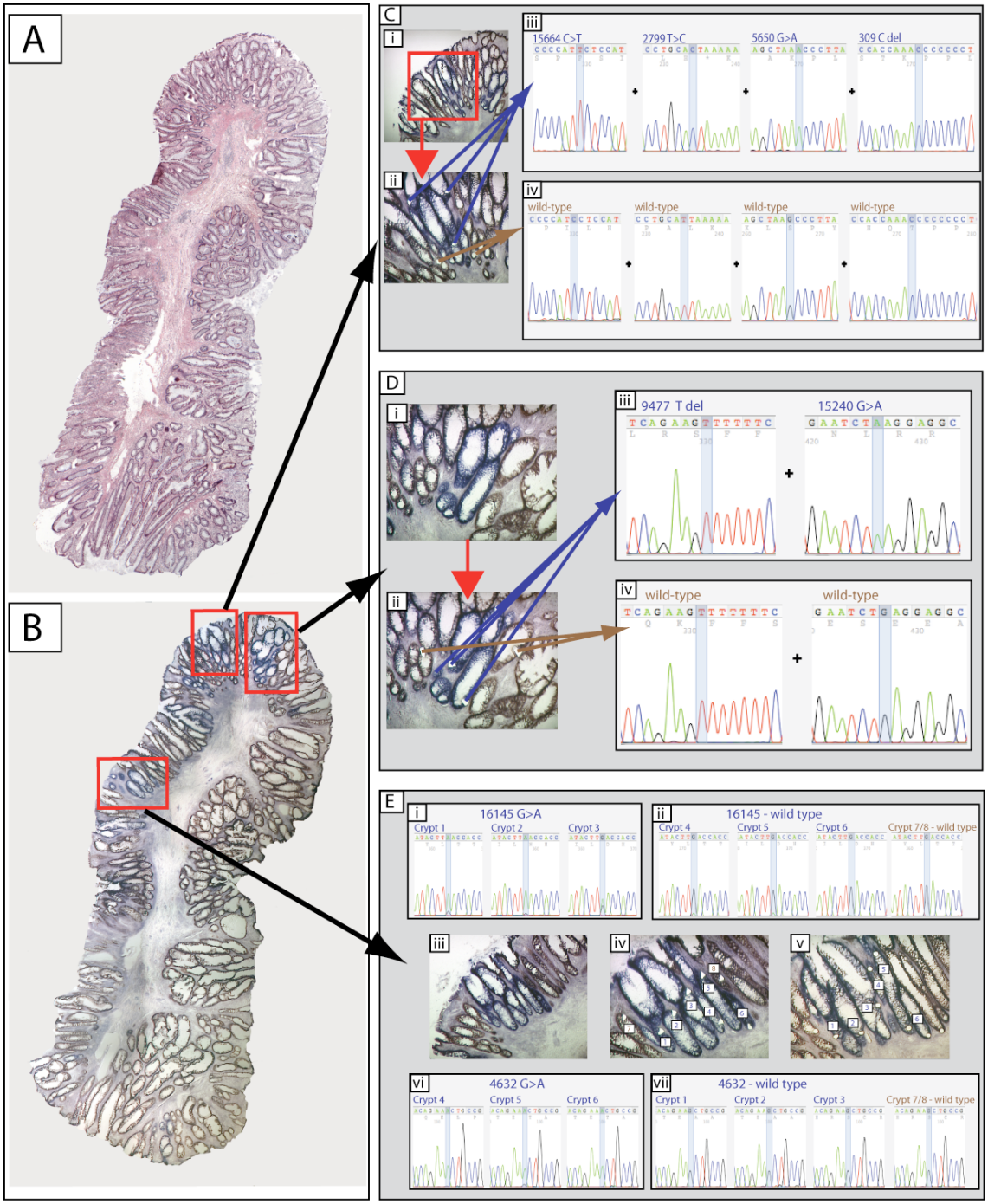

Figure 6.4 Clonally distinct patches within a hyperplastic polyp
Hyperplastic polyp from patient 10 stained for H&E **(A)** and CCO-histochemistry **(B)**. 3 patches of CCO-deficient crypts**: (C)**, **(D)** and **(E)** are highlighted by the red boxes. **(C):** Pre **(i)** and post **(ii)** LCMD images; all blue crypts within this patch harboured four clonal mutations in their mtDNA **(iii-iv):** 15664 C>T in gene *MT-CYB*; 2799 T>C in gene *MT-RNR2*; 5650 G>A in gene *MT-TA* (reverse strand); 309 C del (hyper-variable segment). **(D):** Pre **(i)** and post **(ii)** LCMD images; blue crypts within this patch contained two clonal point mutations in their mtDNA **(iii-iv):** 9477 T del in gene *MT-CO3*, and 15240 G>A in gene *MT-CYB*. **(E):** Pre **(iii)** and post

(iv-v) LCMD of serial sections; blue crypts 1-3 had a clonal 16145 G>A mutation in the hyper-variable segment at heteroplasmic levels **(i)**; blue crypts 4 & 5 were predominantly wild-type but with a small level of the mutation seen; blue crypt 6 and brown crypts 7 & 8 were completely wild-type **(ii)**; blue crypts 4-6 harboured an additional clonal 4632 G>A mutation in gene *MT-ND2* **(vi)** that was not seen in crypts 1-3 of the patch or adjacent brown crypts **(vii)**.

Patch **(E)** (Figure 6.4E) from the hyperplastic polyp contained two clonal point mutations, but these were present at different heteroplasmic levels within crypts from each side of the patch. Figure 6.4E shows sequencing from 6 blue crypts spread across the patch, a 16145 G>A mutation was found at high heteroplasmic levels in crypt 1, however, progressing across the patch the level of the mutation reduced. In crypt 4 the wild-type was seen at high heteroplasmic levels with only a small level of the mutation detected, and the mutation could not be detected in crypt 6. Interestingly, the CCO-deficient crypts 4-6 harboured a separate, homoplasmic clonal mutation within their mtDNA (4632 G>A) that was not seen in any of the other crypts within the patch.

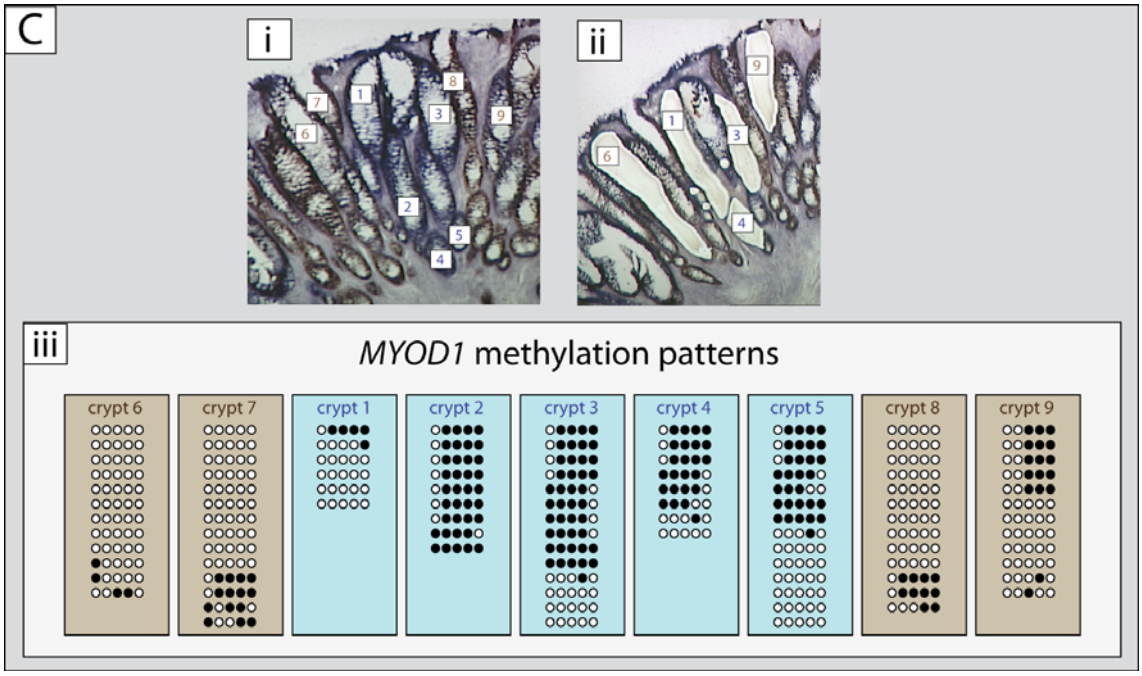

Figure 6.5 *MYOD1* methylation patterns for patch (C) from polyp 1 shown in Figure 6.4

Pre **(i)** and post **(ii)** LCMD pictures (10x). Corresponding methylation patterns at the *MYOD1* locus sampled from blue crypts within the patch (crypts 1-5) and neighbouring wild-type, brown crypts (crypts 6-9) shown in **(iii)**. The methylation patterns at the *MYOD1* locus for crypts within the CCO-deficient sub-clones identified from polyp 1 appeared more similar by their inter-crypt distances when compared to neighbouring, brown, wild-type crypts (see Figure 6.7 for statistical analysis).

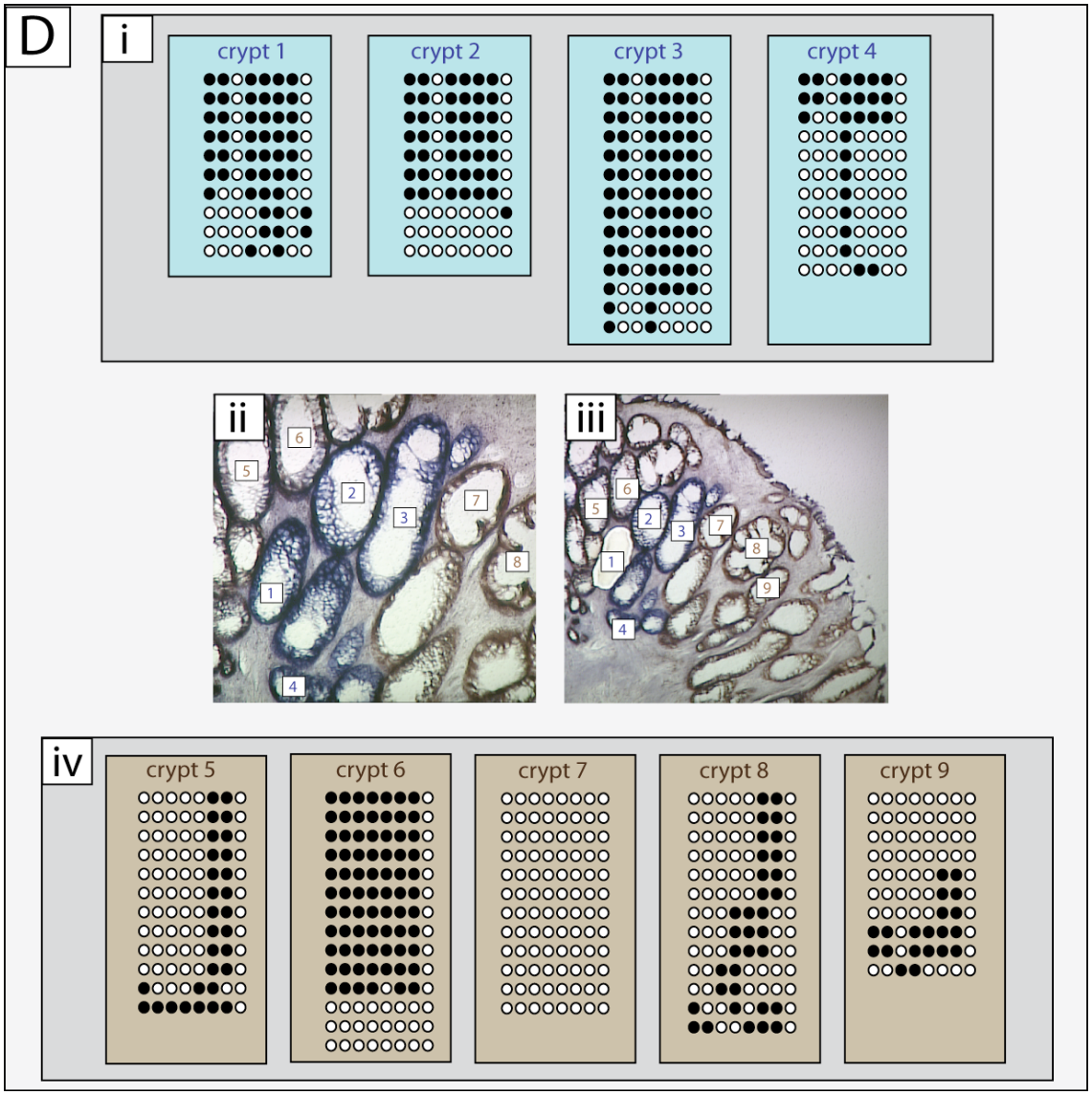

Figure 6.6 *CSX* methylation patterns for patch (D) from polyp 1 shown in Figure 6.4

Pre **(ii)** and post **(iii)** LCMD pictures (10x). Corresponding methylation patterns at the *CSX* locus shown sampled for blue crypts within the patch **(i)** and neighbouring wild-type, brown crypts **(iv)**. The methylation patterns at the *CSX* locus for crypts within the CCO-deficient sub-clones identified from polyp 1 appeared significantly more similar by their inter-crypt distances when compared to neighbouring, brown, wild-type crypts (see Figure 6.7 for statistical analysis)

Methylation patterns at *CSX* and *MYODI* were highly conserved within blue crypts of each of the three clonal, CCO-deficient patches within polyp 1 (Figure 6.5 and Figure 6.6), compared to adjacent, clonally distinct, CCO-proficient crypts. Pairs of crypts within these large, clonally expanded CCO-deficient patches appeared significantly more related to each other by their methylation patterns at the *CSX* locus when compared to adjacent brown, wild-type crypts, by both the inter-crypt distance ($p=5.25 \times 10^{-9}$) and minimum distance ($p=7.52 \times 10^{-7}$) measurements (Figure 6.7A-B). For the *MYOD1* locus there was a similar trend but this did not achieve significance (Figure 6.7C-D): there were less data points collected for the *MYOD1* locus, and it appeared to be a less sensitive marker than *CSX* when studying normal crypts (see Chapter 5), likely due to differences in site-specific methylation rates.

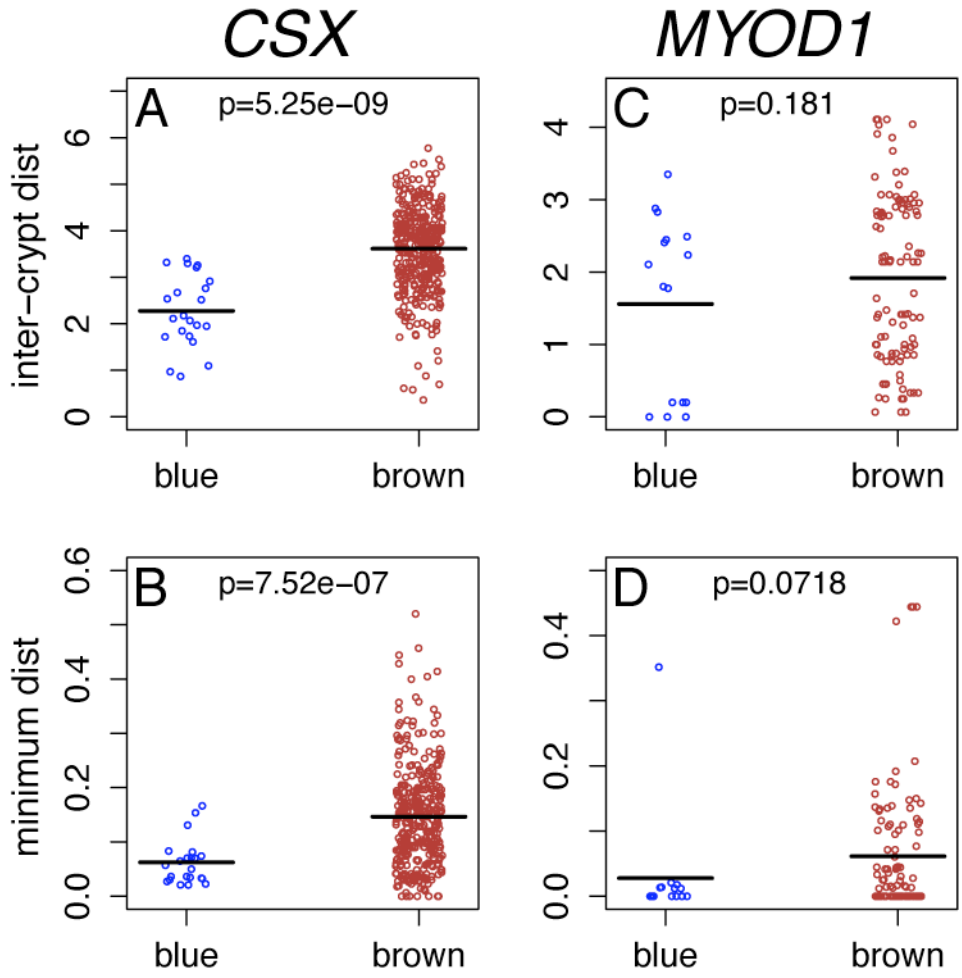

Figure 6.7 Statistical analysis of methylation patterns for all patches within hyperplastic polyp from patient 10 (polyp 1)
Blue dots represent comparisons of two clonally related crypts within a patch; brown dots are comparisons of pairs of crypts that are clonally distinct by their mtDNA. Crypts within clonal patches appear significantly more related by their methylation patterns at the *CSX* locus by both the inter-crypt distance (**A)** and minimum distance **(B)**. Significance was not achieved for the *MYOD1* locus, although the trend was similar to that seen with *CSX* **(C-D)**. Inter-crypt and minimum distances were compared between the two groups using a Mann Whitney U-test and displayed as a scatter-plot: black lines are median values.

Figure 6.8 and Figure 6.9 show adenoma 2 from patient 12. This was a 2.2cm tubular adenoma from the sigmoid colon, displaying low-grade dysplasia (Figure 6.8A). Mutation screening of this adenoma revealed a point mutation in *APC* (c.3964 G>T) that resulted in a premature stop codon (p.E1322*), no additional mutations were identified and the adenoma was MSS (Table 6.1). Two patches of multiple, blue CCO-deficient crypts were seen after histochemistry of serial sections (Figure 6.8B), each containing at least 10 distinct crypts when followed on serial sections. The two patches were seen to become almost continuous on serial sections. LCMD and mtDNA sequencing of cells from multiple blue crypts and adjacent brown, wild-type crypts demonstrated that all blue crypts shared a common, somatic clonal point mutation in their mtDNA (Figure 6.8C-E).

Whole blue and brown crypts were then LCMD for further mutation and methylation analysis (Figure 6.9A). Sequencing confirmed the *APC* mutation was present in all crypts sampled, and was thus clonal for the whole adenoma, and that all crypts were *KRAS* wild-type (Figure 6.9B). The methylation patterns at the *CSX* locus for the adenomatous blue crypts within the patches were similar and conserved, being distinct from those of random, neighbouring brown crypts within the adenoma (Figure 6.9C). Statistical analysis confirmed this, with pairs of adenomatous, clonally-derived blue crypts appearing significantly more related by their methylation patterns when assessing both the inter-crypt distance ($p=1.17 \times 10^{-17}$) and minimum distance ($p=0.005$) (Figure 6.10A-B).

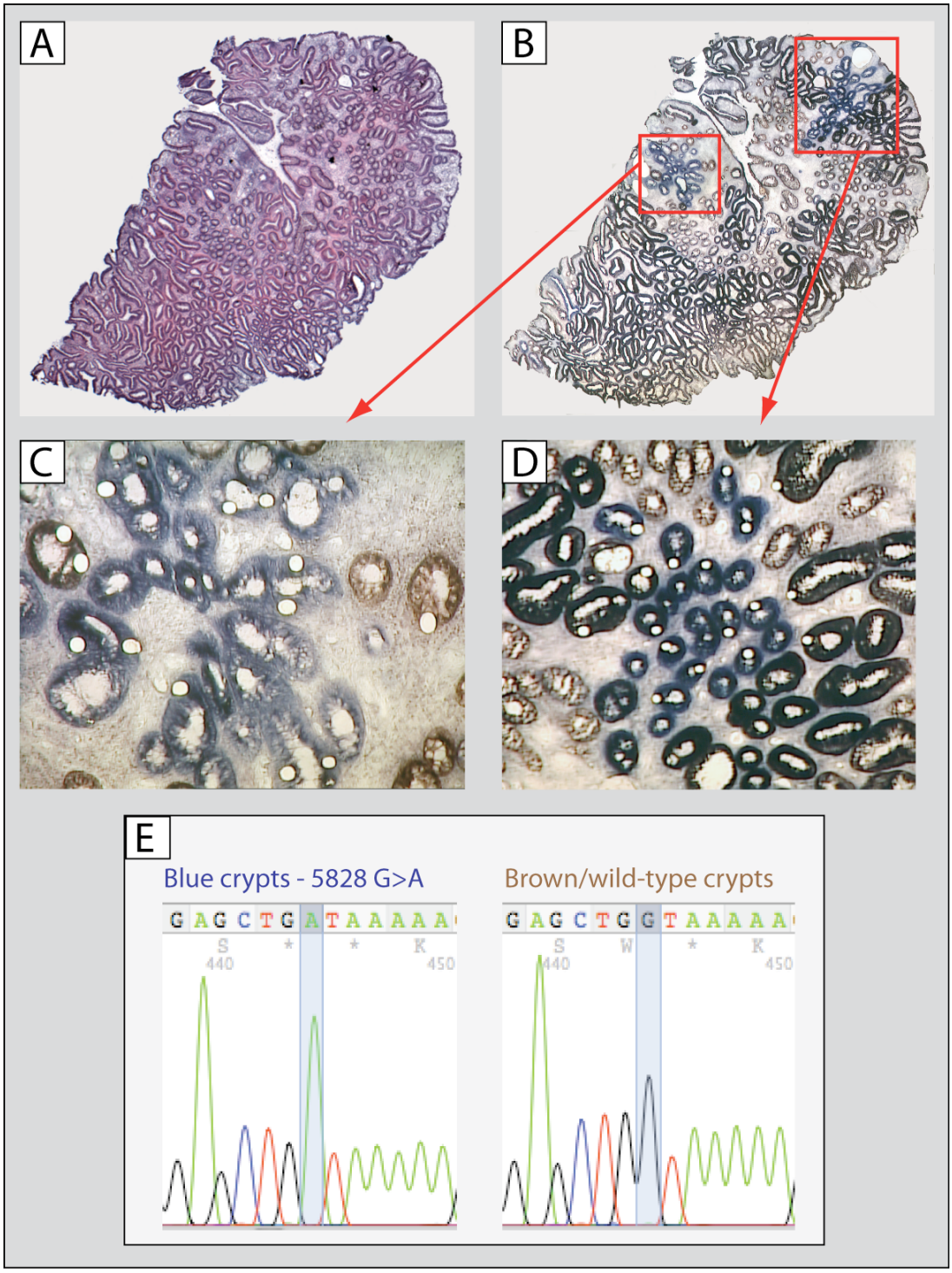

Figure 6.8 Large, clonal patch of CCO-deficient crypts within a human adenoma (patient 12)
(A): Haematoxylin & eosin (H&E) stain showing a tubular adenoma with low-grade dysplasia (adenoma 2). CCO-histochemistry identified two patches of multiple, blue, CCO-deficient crypts **(B)**. LCMD of blue and brown crypts within both patches **(C-D)** demonstrated that all blue crypts contained the same clonal somatic point mutation in their mtDNA that was not seen in adjacent brown crypts **(E)**.

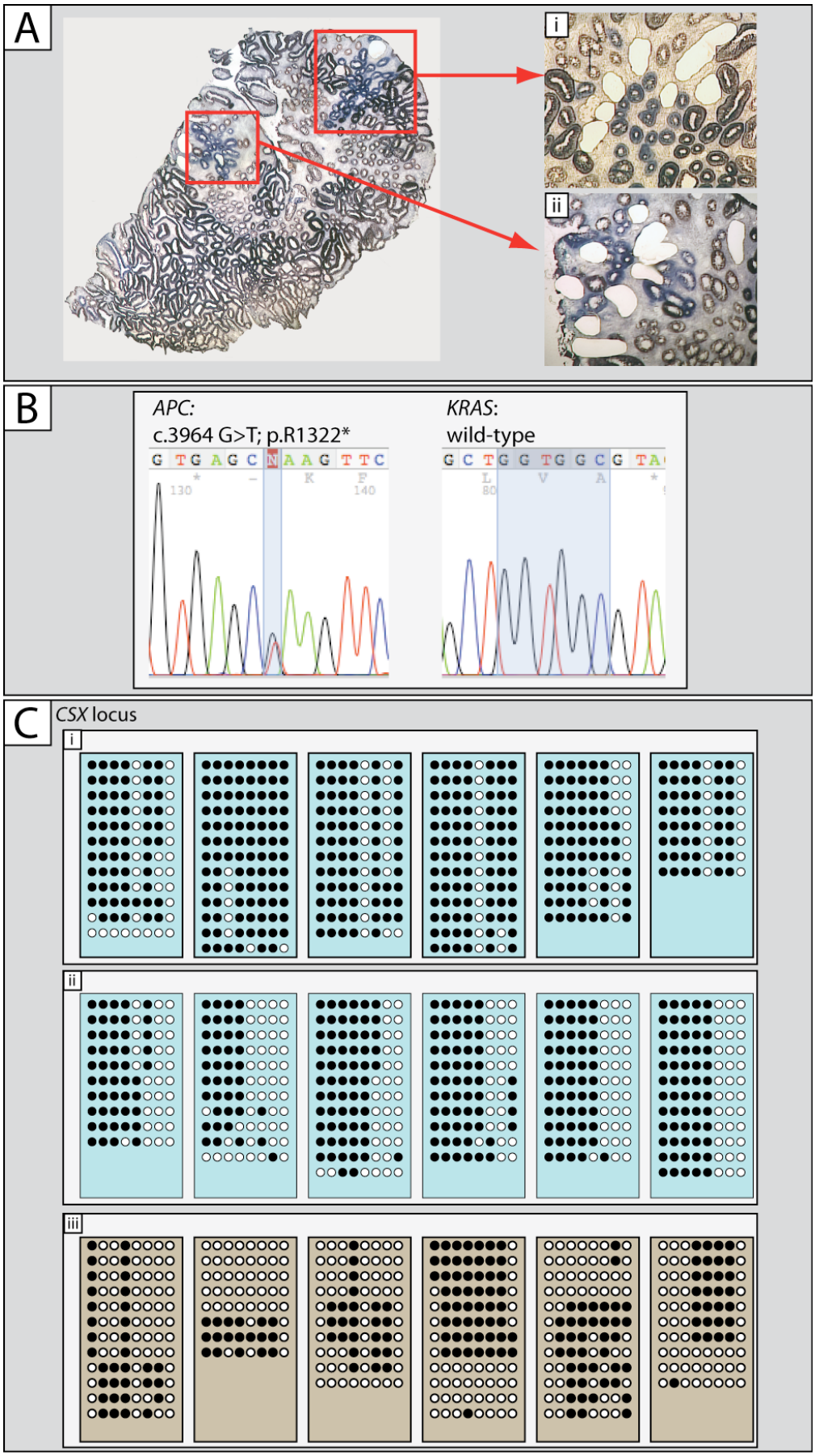
A
i
ii
B
APC:
c.3964 G>T; p.R1322*
KRAS:
wild-type
C
CSX locus
i
ii
iii

Figure 6.9 Methylation patterns at the *CSX* locus for crypts within the clonally expanded patch from adenoma 2
(A): CCO-histochemistry demonstrating the two areas of clonally distinct, CCO-deficient crypts within adenoma 2 from patient 12; whole blue and brown crypts were LCMD from serial sections **(i-ii)**. **(B):** PCR confirmed the *APC* mutation detected at screening in all blue and brown crypts (c.3964G>T); all crypts were *KRAS* wild-type. **(C):** Methylation patterns at the *CSX* locus appeared relatively conserved for all blue crypts **(i-ii)**, and distinct from randomly sampled brown crypts **(iii)**. See Figure 6.10 for statistical analysis.

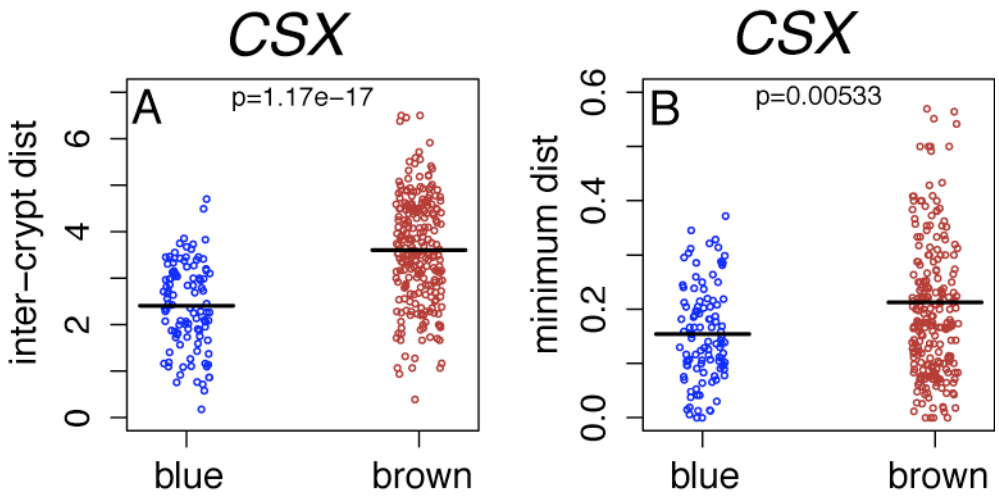

Figure 6.10 Statistical comparison of methylation patterns for blue versus brown crypts within adenoma 2
Blue dots represent pairs of adenomatous crypts within the CCO-deficient patch shown to be clonally related by a somatic mutation in their mtDNA. Brown dots represent pairs of adenomatous crypts that were wild-type by their mtDNA. Blue crypts within the patch appear significantly more related by both the inter-crypt distance **(A)** and minimum distance **(B)** with p-values of 1.17 $x10^{17}$ and 0.005 respectively. Inter-crypt and minimum distances were compared between the two groups using a Mann Whitney U-test and displayed as a scatter-plot: black lines are median values.

6.4 Adenomas clonal for *APC* and *KRAS* mutations demonstrate significantly less methylation pattern diversity

Whole crypts, spaced at regular intervals throughout adenomas 1, 3, 4, 5 and 6 were LCMD from serial sections and their methylation patterns at the *CSX* and *MYOD1* locus determined. For each crypt micro-dissected, the genomic mutations detected at screening were confirmed by PCR before taking forward for methylation analysis.

Figure 6.11 shows the 2.2cm tubulo-villous adenoma (TVA) with low-grade dysplasia from patient 11 (adenoma 1), and Figure 6.12 the 1.3cm TVA from patient 13 (adenoma 3). Mutation screening revealed both adenomas had mutations in *APC* and *KRAS*, and this was confirmed in all of the individual crypts micro-dissected for methylation analysis (Figure 6.11A & 6.12A). Methylation patterns at the *CSX* and *MYOD1* loci were relatively conserved between crypts when compared to adenomas with mutations in *APC* only (Figure 6.15), showing little diversity across the whole adenoma in both cases (Figure 6.11C & 6.12B).

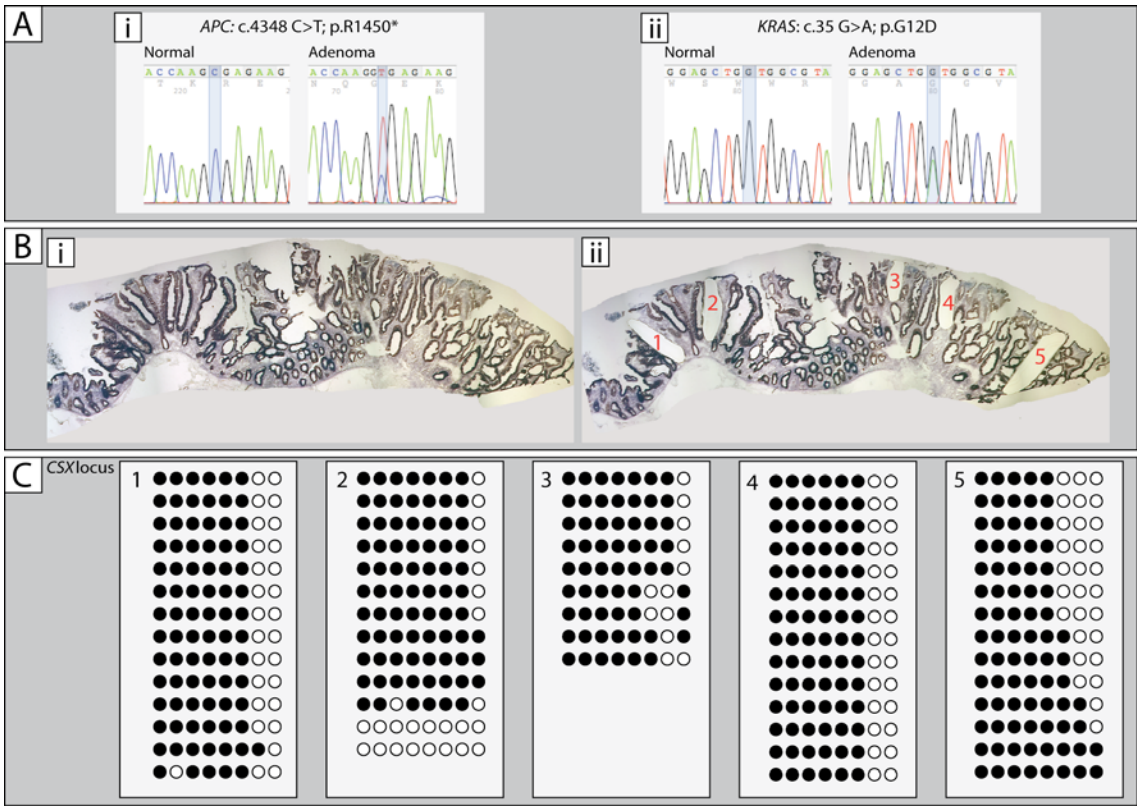

Figure 6.11 Methylation patterns at the *CSX* locus for adenoma 1
(A): PCR and sequencing shown of individual crypts within the adenoma and adjacent normal tissue; All adenomatous crypts contained a clonal point mutation in *APC* (c.4348 C>T) **(i)** and a clonal p.G12D *KRAS* mutation **(ii)** whereas normal crypts were wild-type. **(B):** Pre **(i)** and post **(ii)** LCMD of 5 crypts spaced across the section of the adenoma. **(C):** Methylation patterns at the *CSX* locus for each corresponding crypt numbered in **(B)**.

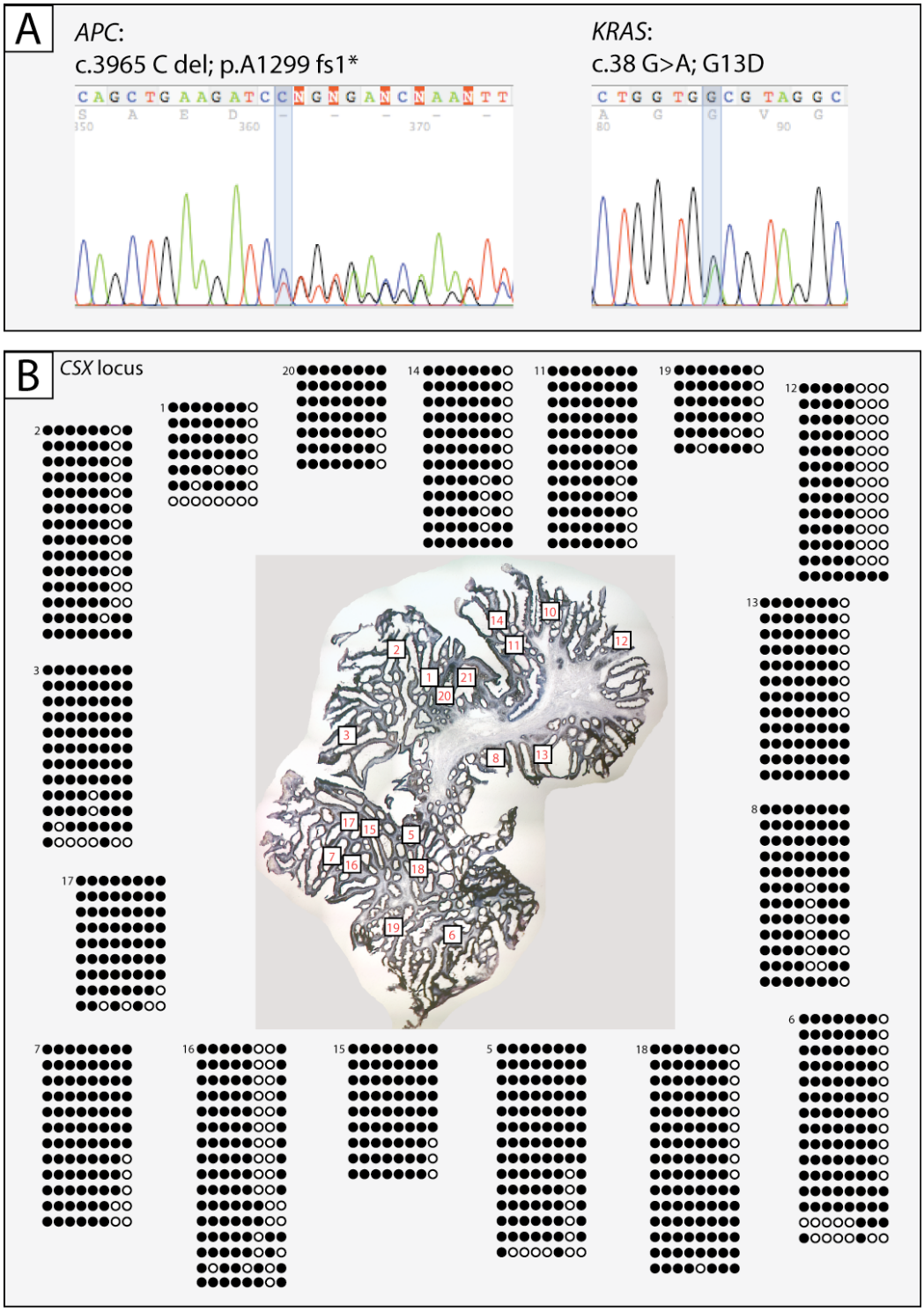

Figure 6.12 Methylation patterns at the *CSX* locus for adenoma 3
(A): PCR and sequencing shown of individual crypts within the adenoma: All adenomatous crypts contained a clonal deletion in *APC* (c.3965 C del) and a clonal G13D *KRAS* mutation **(B):** Methylation patterns at the *CSX* locus for crypts LCMD from across the adenoma; numbered methylation tags relate to the corresponding numbered crypt shown in the overview section.

Adenomas 2, 4 and 5 had mutations in the *APC* gene but were otherwise wild-type for *KRAS, BRAF* and *TP53*, MSS and there was no evidence of LOH (Table 6.1). In contrast to the adenomas harbouring *KRAS* mutations, methylation patterns between crypts within these adenomas demonstrated significantly more variance (Figure 6.15). Repeat PCR for all individual crypts micro-dissected for methylation analysis from adenomas 2, 4 and 5 was performed to confirm they were indeed clonal for the *APC* mutation found at screening, and were all wild-type for *KRAS*. Figure 6.13 and Figure 6.14 show the methylation patterns at the *CSX* locus for crypts from adenoma 2 and adenoma 5 respectively.

The inter-crypt and minimum distances – the measurements of methylation pattern diversity between crypts within each adenoma - was calculated for all crypts sampled within adenomas 1, 2, 3, 4 and 5. The data for adenomas 1 and 3 was then compared to adenomas 2, 4, and 5 (Figure 6.15A-D). Crypts within adenomas clonal for *APC* and *KRAS* mutations demonstrated significantly less inter-crypt diversity, and thus appeared significantly more related by their methylation patterns than crypts with adenomas clonal only for an *APC* mutation at the *CSX* locus ($p=4 \times 10^{-26}$ for inter-crypt distance and $p=1.8 \times 10^{-25}$ for minimum distance) (Figure 6.15A-B). When assessing the *MYOD1* locus, the same trend was seen for the inter-crypt distance ($p=0.04$), although no significant difference between the adenoma types was detected when using the minimum distance (Figure 6.415C-D).

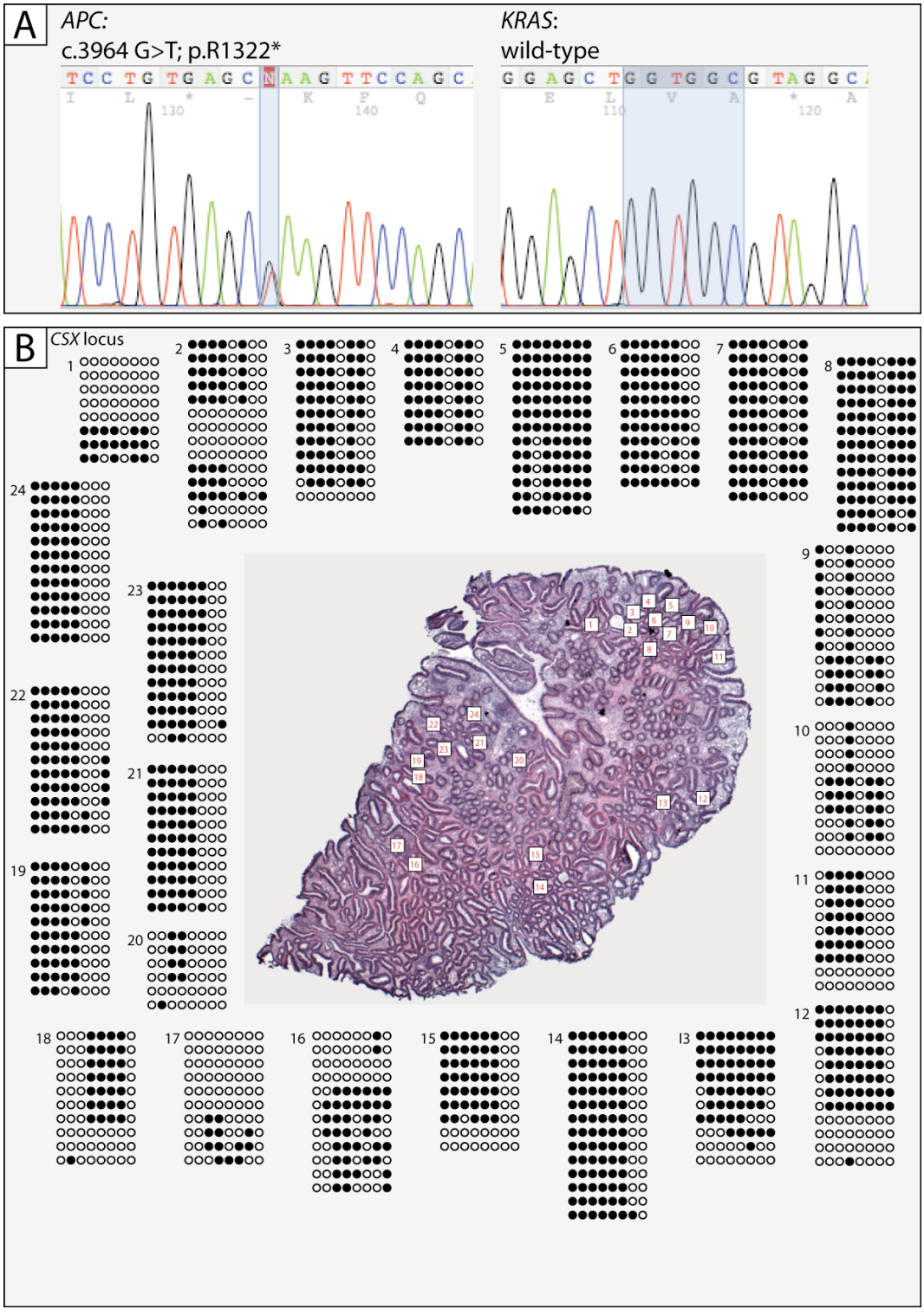

Figure 6.13 Methylation patterns at the *CSX* locus for adenoma 2
(A): PCR and sequencing of individual crypts from within the adenoma: All adenomatous crypts contained a clonal point mutation in *APC* (c.3964 G>T) and were wild-type for *KRAS*. **(B):** Methylation at the *CSX* locus for crypts LCMD from across the adenoma; numbered methylation tags relate to the corresponding numbered crypt shown in the overview section.

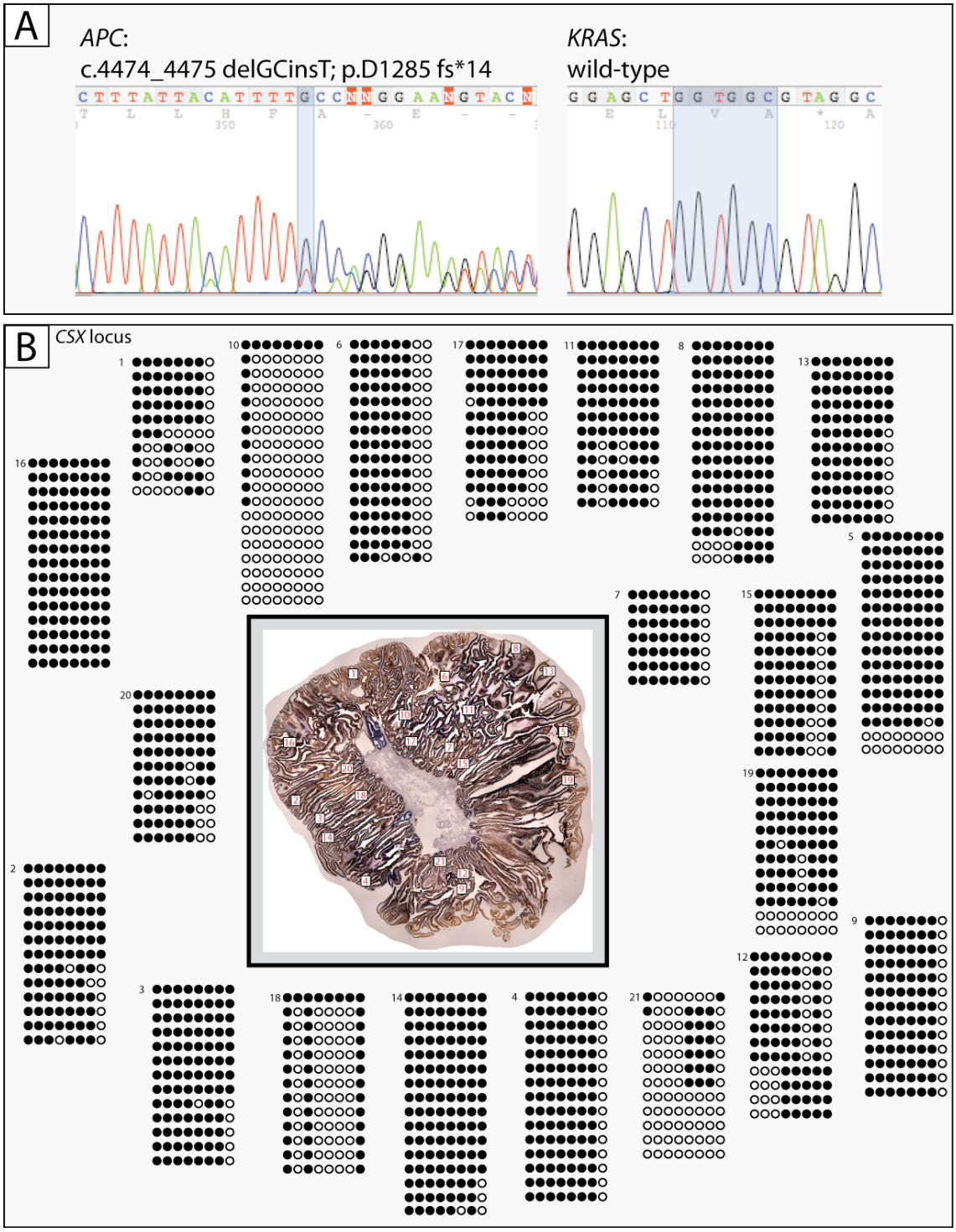

Figure 6.14 Methylation patterns at the *CSX* locus for adenoma 5
(A): PCR and sequencing of individual crypts from within the adenoma: All adenomatous crypts contained a clonal indel in *APC* (c.4474_4475 GCdelTins) and were wild-type for *KRAS* **(B):** Methylation patterns at the *CSX* locus for crypts LCMD from across the adenoma; numbered methylation tags relate to the corresponding numbered crypt shown in the overview section.

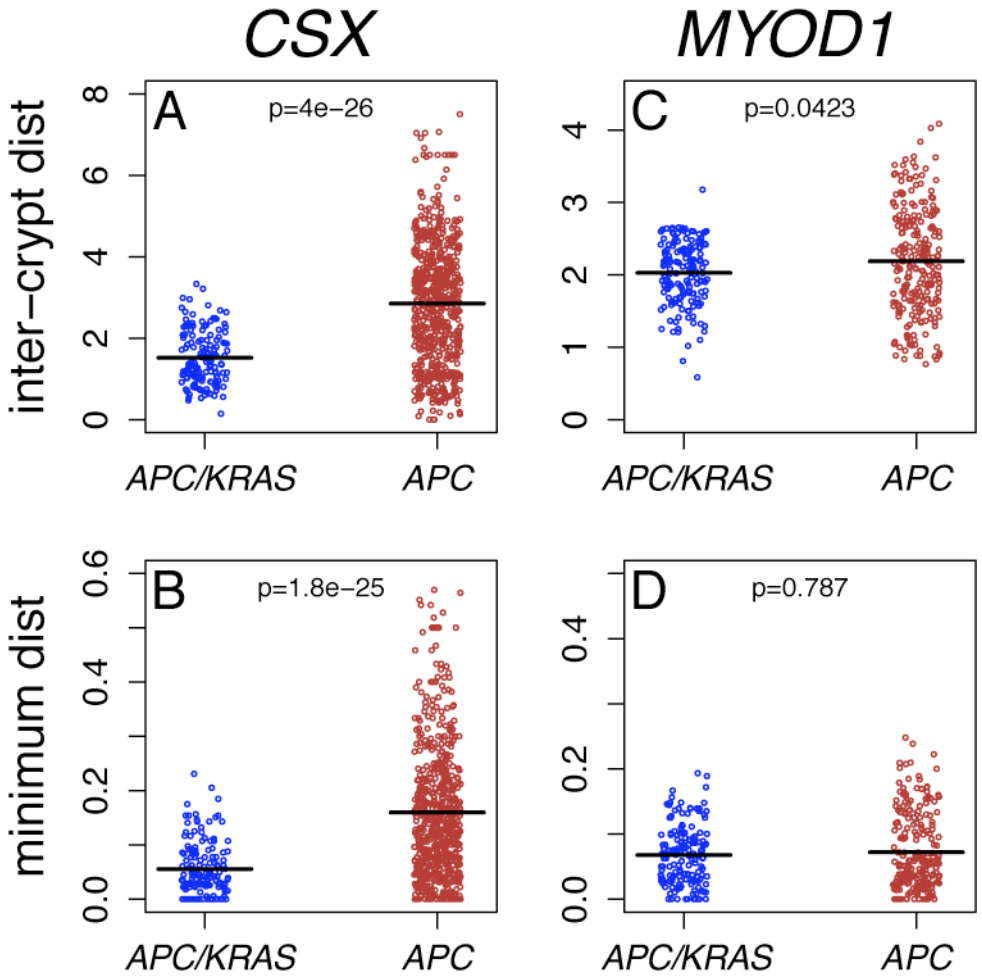

Figure 6.15 Crypts within adenomas that are clonal for *APC* and *KRAS* mutations appear significantly more related by their methylation patterns
Blue dots are values for pairs of crypts within adenomas 1 and 3, where all crypts contained clonal *APC* and *KRAS* mutations; brown dots represent comparisons of pairs of crypts within adenomas clonal only for an *APC* mutation (adenomas 2, 4 and 5). Crypts from *APC/KRAS* adenomas appeared significantly more related by both the inter-crypt **(A)** and minimum distance **(B)** at the *CSX* locus. The same trend was seen at the *MYOD1* locus for the inter-crypt distance **(C)**, but no significant difference was detected when comparing crypts by the minimum distance **(D)**. Inter-crypt and minimum distances were compared between the two groups using a Mann Whitney U-test and displayed as a scatter-plot: black lines are median values.

6.5 Adenomas appear to display punctuated growth

The methylation data across the adenomas was analysed in order to infer patterns of overall growth, by correlating the physical distances of crypts within an adenoma to their respective epigenetic distances, as measured by the inter-crypt distance at methylation tags of the *CSX* locus. Each crypt within adenomas 1, 3 and 5 was ranked according to the physical number of crypts separating it from all the other crypts within the lesion (physical distance), this was then plotted against the respective ranked epigenetic distance for the same pairs of crypts. If the lesions had grown in a constant, linear fashion the physical distances would be expected to correlate with the epigenetic distances between pairs of crypts over short length scales. However, for the adenomas analysed, no obvious correlation was seen (Figure 6.16), suggesting that there had been punctuated growth: Methylation patterns of crypts across the whole adenoma would appear equally similar after a period of rapid expansion, and tags of crypts in close spatial proximity would diverge during quiescent stages.

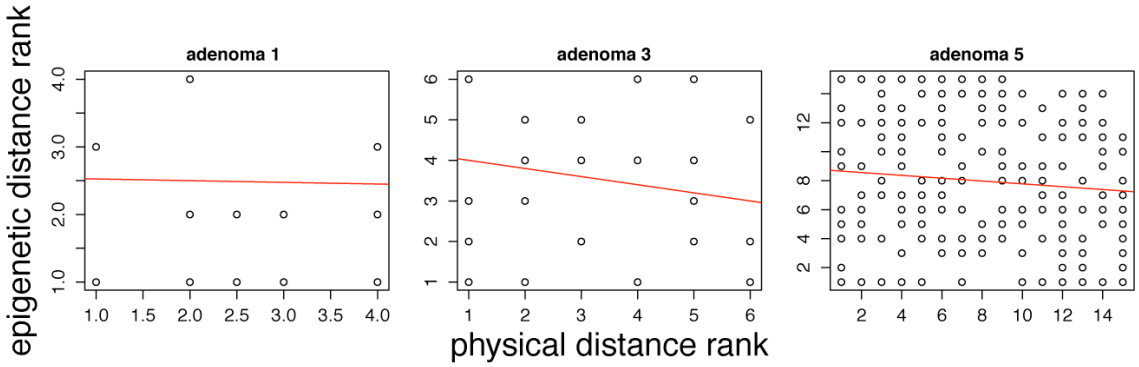

Figure 6.16 The physical separation of crypts within adenomas is not reflected by their epigenetic distances
Crypts within adenomas 1, 3 and 5 that had been micro-dissected to assess methylation tags at the *CSX* locus were ranked according to the number of crypts separating them, this was then plotted against the rank of their respective epigenetic distances. Trend lines are shown in red: p-values for the gradient of the trend line being different to 0 were non-significant for all adenomas.

6.6 *KRAS* sub-clones can be identified and appear to be expanding rapidly

On screening adenoma 6 from macro-dissected tissue, a small peak of mutant *KRAS* was identified (c.35G>A), in addition to a mutation in *APC* (c.3925_3929delGAAAA). Subsequent LCMD of individual crypts from across the adenoma identified two small, clonally distinct areas of adenomatous crypts that harboured the *KRAS* and *APC* mutations, whereas remaining crypts where clonal only for *APC* and *KRAS* wild-type (Figure 6.17A-B). Subsequent methylation analysis was unfortunately only successful for two crypts from the *KRAS* sub-clone, however this did show them to have very similar *CSX* methylation patterns, whereas methylation tags were quite divergent amongst randomly selected *KRAS* wild-type crypts (Figure 6.17B).

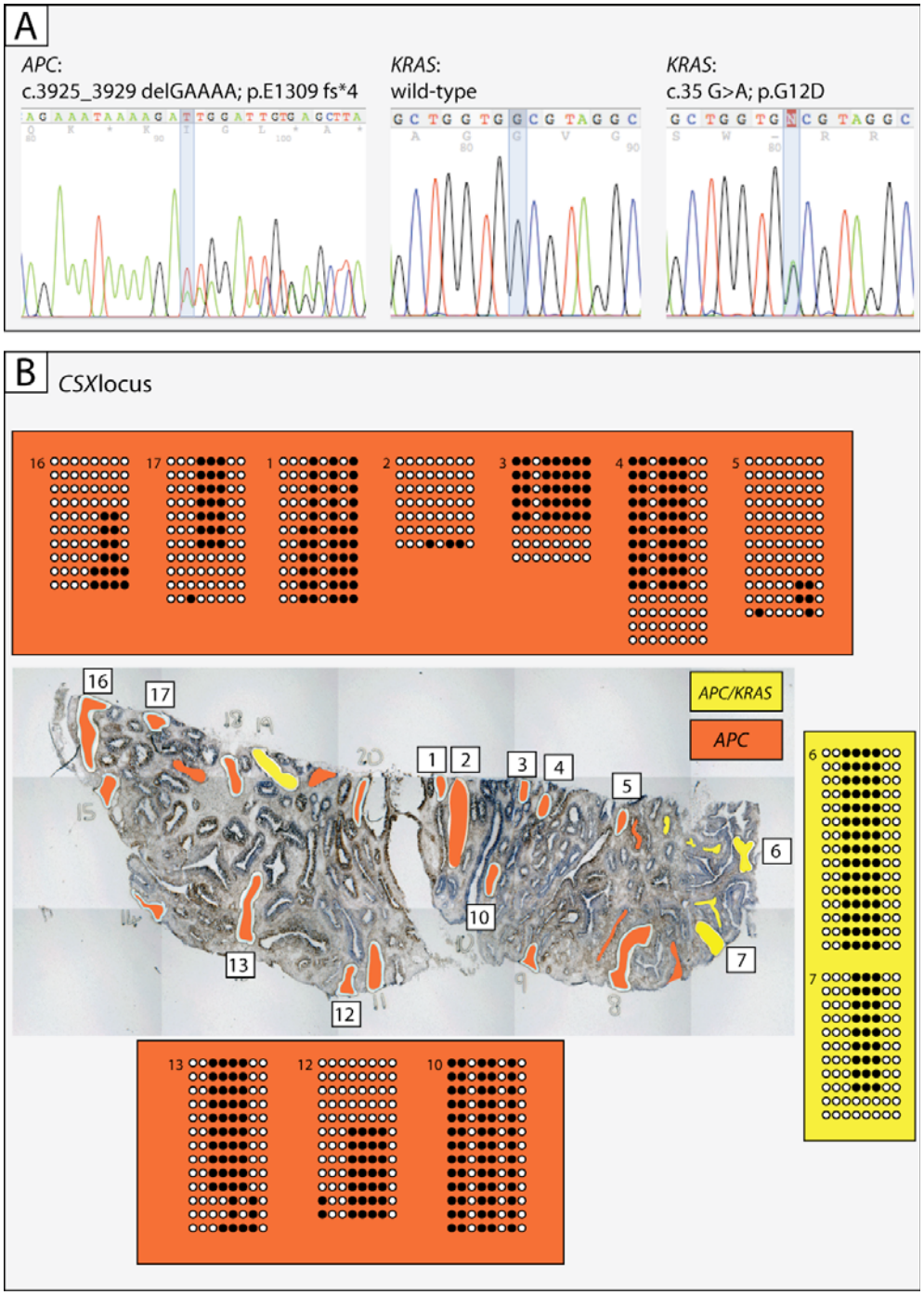

Figure 6.17 Methylation patterns at the CSX locus for the KRAS sub-clone within adenoma 6
(A): All micro-dissected crypts contained a mutation in *APC* (c.3925_3929 delGAAAA); most crypts (highlighted in red in **(B)**) were wild-type for *KRAS*, however 6 crypts had an additional *KRAS* mutation (highlighted in yellow in **(B)**).
(B): Methylation tags for crypts at the *CSX* locus; limited data from two *KRAS* mutated crypts showed them to have very similar methylation patterns; numbered methylation tags relate to the corresponding numbered crypt shown in the overview section.

6.7 Discussion

Crypt fission has been shown to be the major mechanism by which early adenomas grow [93], so-called *bottom-up* spread, with dysplastic cells invading adjacent normal crypts occurring later on in adenoma development – *top-down* spread. Wong et al [94] demonstrated that crypt fission events are significantly more frequent in both hyperplastic polyps and adenomas compared to normal tissue, with an upward expansion of the proliferative compartment in adenomas. However, there is limited data on how adenomas grow, with longitudinal barium and endoscopic studies showing that many smaller adenomas appear to be static over many years. This project aimed to identify clonal sub-populations within polyps and adenomas in order to investigate growth dynamics by analysing methylation patters at CpG islands of non-expressed genes.

All of the sporadic adenomas were greater than 1cm in size (measured post-resection by a pathologist) and 88% (8/9) had mutations in the *APC* MCR, higher than that reported in other studies [84, 133]. The tissue used in this study was fresh frozen which enables much more efficient PCR of DNA, almost all other studies that have looked at the frequency of *APC* mutations in sporadic adenomas have used FFPE tissue, a fixing process that fragments DNA and makes PCR of the entire MCR much more challenging. Thus mutations may have been missed in previous studies due to incomplete amplification of the MCR of *APC*. It has been reported that in FAP patients, due to the differing functional consequences of mutations in the MCR of *APC*, germ-line mutations around codon 1296 or 1309 are highly selective for LOH in the other allele [143]. Of the nine adenomas screened, one had a mutation in codon 1299 and two in codon 1309 of *APC* and surprisingly none had evidence of LOH by the microsatellites used here; however this does not exclude LOH, since focal chromosomal losses may not be detected by LOH markers. In this study 33% (3/9) of the adenomas screened were found to have mutations in the *KRAS* oncogene, which is lower than that reported by Vogelstein [113] (58%). However, the incidence of *KRAS* mutations in colorectal cancer does vary from 20-50% between studies [137, 170]. The absence of mutations in *TP53* and

17p LOH is consisted with previous work associating these with more advanced (high-grade dysplasia) adenomas and the progression to cancer [113, 114, 144]. None of the adenomas had evidence of MSI or contained mutations in the *BRAF* oncogene, suggesting that there was no CIMP in these lesions, and no difference in global genomic DNA methylation level as measured by LINE-1 Pyrosequencing® was seen between normal and adenomatous crypts.

Multiple, large, clonally distinct CCO-deficient patches of crypts were seen within a hyperplastic polyp. Crypts within these patches appeared significantly more related by their methylation patterns compared to random brown, wild-type crypts from within the same lesion. This is in contrast to what was seen in normal tissue (see Chapter 5), and thus these crypts must have expanded rapidly by fission in order for the methylation patterns to remain similar over such large patches of crypts. Similarly, a large CCO-deficient sub-clone was identified within an adenoma: All crypts within the adenoma harboured the same, clonal mutation in *APC*, however the CCO-deficient crypts were clonally distinct by sharing a common somatic mutation in their mtDNA. Analysis of the methylation patterns again suggested that the CCO-deficient sub-clone had expanded rapidly within the adenoma. The reason for the CCO-deficient sub-clone having a selective advantage in this case is not clear, as no additional tumorigenic mutations other then the *APC* mutation were detected within the adenoma at screening. However, there are likely many more tumorigenic changes occurring in adenomas that were not screened for in this study, and LOH analysis on a crypt-by crypt basis might reveal a selective advantage.

Methylation data from crypts micro-dissected across adenomas with a clonal *APC* mutation were compared to adenomas with a clonal *APC* and *KRAS* mutation (*APC/KRAS)* in all crypts; pairs of crypts within *APC/KRAS* adenomas appeared significantly more related by their both their inter-crypt distance and minimum distance. As adenomas grow the time to the most recent common ancestor between crypts increases and individual crypt methylation patterns will diverge, thus the *APC/KRAS*

adenomas appeared a much younger population by their methylation patterns, suggesting they had not yet had time to diversify, consistent with their being a recent selective sweep and rapid clonal expansion in these lesions. It is not possible to say whether the *KRAS* or *APC* mutation initiated this, as both mutations were found in all crypts. These observations could have been tested further by immunohistochemical staining for markers of cellular proliferation such as KI-67 and GEMININ expression, in addition to assessing levels of apoptosis using CASPASE-3 expression or TUNEL (terminal deoxynucleotidyl dUTP transferase nick end labelling) assays. Comparing the spatial distances of crypts within adenomas to their respective epigenetic distances, as measure by the inter-crypt distance, suggested that growth had not occurred in a constant linear fashion and was consisted with a punctuated growth pattern, thus periods of rapid clonal expansion may be followed by periods of relative quiescence.

There is evidence for significant interaction between the Wnt and MAPK pathways: Li et al [185] showed that expression of oncogenic *KRAS* mutations stimulates Wnt signalling in colon cancer cell lines, and compound transgenic mouse models have demonstrated that expression of oncogenic *KRAS*, in addition to *APC* loss, promotes adenoma and tumour progression in the intestines and further up-regulates Wnt/β-catenin signalling pathways [186, 187]. The *KRAS* sub-clone identified within adenoma 6 does allow clonal ordering however, and the initial methylation data suggests this is a rapidly expanding population, providing potential direct evidence for a growth advantage provided by a *KRAS* mutation in addition to *APC* in human adenomas for the first time. Further work examining the dynamics of *KRAS* sub-clones within adenomas will allow more substantiated conclusions and relative growth rates to be estimated.

It was demonstrated in Chapter 5 that *CSX* methylation tags appeared to record clonal ancestry more effectively than *MYOD1* tags: clonally-derived CCO-deficient crypts had significantly more similar *CSX* methylation patterns than unrelated crypts (Figure 5.9), and *CSX* methylation pattern diversity scaled with clone size in partially-mutated crypts (Figure 5.7), whereas these trends were not consistently observed for *MYOD1*. The

methylation patterns at the *BGN* locus were more similar for unrelated crypts than those derived by fission (Figure 5.9), suggesting it is not able to accurately record clonal ancestry over long time periods. This differential ability to record ancestry is likely to result from both the locus and CpG-site specific (de-)methylation rates. The same trends for both *CSX* and *MYOD1* were observed when statistically analysing the methylation patterns for crypts within adenomas, supporting the validity of the findings presented in this chapter, however no data is presented here for the *BGN* locus as, for reasons that were not clear, it was not possible to amplify this locus efficiently enough from the adenoma tissue in order to sample the methylation status.

It is widely accepted that tumorigenesis is a form of somatic evolution, and it has been suggested that selection, without increased mutation rates, is sufficient to explain the initiation and expansion of colorectal adenomas [138, 142]. Although adenomas less than 1cm in size do not appear to be actively growing when assessing their size, dynamic clonal expansion might go unnoticed if any increase in crypt number is balanced by crypt death [188] – a hypothesis that could be tested by looking at apoptotic indices in sub-clone populations. The results presented here provide evidence that sub-clones within adenomas and polyps can be identified that are rapidly expanding by fission compared to neighbouring crypts, and *APC* mutated adenomas with oncogenic *KRAS* mutations appear to have undergone recent clonal expansions. In addition, adenomas appear to have periods of little or no growth that enable the methylation patterns of adjacent crypts to diverge. Thus, adenoma growth may be characterised by cycles of rapid clonal expansion as new, selectively advantageous mutations are acquired, followed by periods of little or no growth; during the apparent static phases there may be dynamic competition between sub-clones that drives the further evolution of the lesion. However, there are several limitations of the data set presented in this chapter: the number of adenomas studied so far is small, and further work needs to be done to validate the findings presented here, in addition growth dynamics are being inferred from a snapshot in time - an experimental animal model could be developed that would allow the hypotheses generated to be tested further under more controlled conditions with the collection of temporal data.

Chapter 7. Conclusions

7.1 Delineating stem cell clones in normal human intestine

The results presented here have enabled, for the first time in human tissue, the identification of a single stem cell lineage and their direct progeny within human colon crypts. There appear to be at least 2 stem cell lineages within the human colonic niche, located at the base of the colon crypt and around the +4 position in the small bowel. Mapping of clones has allowed the visualisation of niche succession and clonal expansion within human intestinal crypts.

7.2 Niche succession in the normal human colon crypt occurs over many years

Analysis of the methylation patterns of CCO-deficient stem cell clones occupying different areas of the crypt, and thus at different stages of niche succession, suggests that this is slow process, taking many years in the human colon.

7.3 There are at least 10 years between crypt fission cycles in normal human colon

Patches of CCO-deficient crypts, known to be clonally-derived by fission events, appear disparate by their methylation patterns, thus there is sufficient time (years) between fission events to allow divergence. Mathematical modelling suggests that approximately 10 years is the lower limit for crypt fission events in normal human colon.

7.4 Human colorectal adenomas may exhibit punctuated growth, with clonal competition driving the evolution of these lesions

Genetically distinct sub-clones have been identified within human adenomas and polyps that are expanding rapidly compared to surrounding crypts within the lesion. Adenomas with mutations in *APC* and *KRAS* appear to have undergone recent clonal expansions, and there is evidence that oncogenic *KRAS* expression on the background of *APC* mutations offers a growth advantage in human adenomas. Analysis of methylation patterns within adenomas suggests they may display punctuated growth, with cycles of rapid clonal expansion followed by relative quiescence.

Chapter 8. Summary Discussion

Much of the current knowledge on the location of the intestinal stem cell niche, stem cell number and dynamics of clonal expansion within crypts is based on mutagenic studies [10, 17, 70, 71] and, more recently, genetic lineage tracing experiments in transgenic mice [19, 20, 29], employing techniques that are impractical in humans. No experimental evidence has yet conclusively demonstrated the stem cell niche in the human intestine. The 'gold standard' criteria for identifying tissue-specific stem cells is now established as: (i) self-renewal and (ii) multi-lineage potential. Thus, the aims of Chapters 3 and 4 were to use naturally occurring clonal somatic point mutations in mtDNA to identify and map, for the first time, the progeny of a putative single stem cell lineage within human crypts, and then to demonstrate the multi-lineage potential of these clones, thus providing evidence that they are, indeed, stem cell populations.

Why is stem cell biology important? It has long been proposed that cancer is a consequence of tumorigenic mutations accruing in a single, tissue-specific stem cell [115], based mainly on hypothetical reasoning that these are the only cell lineages with sufficient life-span to accumulate the multiple mutations required for tumorigenesis. However, there is now experimental evidence to support stem cells as the cells of origin in colorectal carcinogenesis: Knocking out *Apc* in Lgr5+ stem cells results in rapid adenoma formation, whereas knocking out *Apc* in non-stem cells does not [176]. Furthermore, a recent study has demonstrated that non-steroidal anti-inflammatory drugs (NSAIDS) are able to target stem cells that have accumulated tumorigenic mutations, and can eliminate them by the induction of apoptosis [189]: In an *Apc* $^{Min/+}$ mouse model, dietary sulindac (an NSAID) induced apoptosis in intestinal stem cells with nuclear or phosphorylated β-catenin, and in human colonic polyps NSAIDS were shown to induce apoptosis in cells with aberrant Wnt signalling [189]. Thus, understanding the stem cell biology of the human intestine, and the key mechanisms that are dysregulated in the

initiation of dysplasia, may enable the development of effective chemoprevention by targeting aberrant stem cell populations. Indeed, studies have consistently demonstrated that the use of aspirin [190-193] and cyclo-oxygenase-2 enzyme (COX-2) inhibitors [194, 195] is associated with around a 20% reduction in adenoma recurrence in patients. A recent meta-analysis pooling long-term follow up data from four randomized, double-blind, placebo-controlled trials of aspirin treatment [196] demonstrated that regular aspirin significantly reduced long-term risk of colon cancer – interestingly, this reduction was almost entirely due to a decrease in the incidence of proximal colorectal cancers. The biology of the stem cell niche has long been shown to differ with progression from the small bowel to the rectum, with the proximal colon itself appearing more like the small intestine [27, 32, 33], thus it is interesting to wonder whether the apparent chemo-preventative effects of aspirin confined to the proximal human colon are due to specific effects on the stem cells and their niche here, a hypothesis that could be explored in mouse models.

The data presented for the human colon in Chapters 3 and 4 was consistent with the prevailing dogma from murine studies, that there are multiple, related stem cells located at the base of the colonic crypt that undergo successive niche succession cycles. Only a small amount of data was presented for the small intestine, due to difficulty in obtaining fresh frozen tissue, however the limited results did indicate that the niche in the human small bowel is around cell position 4, above the basal Paneth cells. The Barker et al [19] Lgr5 lineage tracing study was a series of seminal experiments, and re-defined the criteria for identifying putative stem cells, providing robust evidence that, as proposed thirty years previously [5, 24], the crypt based columnar cells were a stem cell population in the small bowel. However, the situation in the small bowel niche is still as debated as ever, with further studies employing the techniques used by Barker et al [19] to perform lineage-tracing using other markers such as *Bmi1* demonstrating putative stem cells concentrated around the +4 position [20, 29]. Thus, there does appear to be a heterogeneous stem cell population in the mouse small bowel, with actively cycling stem cells at the base below more quiescent stem cells around cell position 4 – an organisation not too far removed from the hierarchical arrangement of the niche proposed by Potten and colleagues almost

fifteen years ago [27]. In the human small and large intestine, further work combining mapping of CCO-deficient cell populations within crypts with the new panel of putative stem cell markers developed in mouse lineage studies might allow further characterisation of the human niche.

In Chapter 5, the rates of niche succession and crypt fission, the methods by which tumorigenic mutations in stem cells are thought to initially become established within the epithelium, were investigated in normal human colon by studying the methylation patterns at CpG loci of non-expressed genes in CCO-deficient cell populations. Similar techniques were then applied in Chapter 6 to investigate the dynamics and clonal expansion of early human adenoma growth. The methylation patterns of stem cell lineages expanding within the normal crypt at different stages of clonal succession suggested that niche succession can take many years, furthermore crypt fission events in normal human colon appear to be separated by at least 10 years. The situation observed in early human adenomas and polyps was quite the opposite, with large patches of clonally expanded CCO-deficient crypts identified that appeared very similar by their methylation patterns, indicating that the clonal expansion rates are significantly up-regulated to allow the rapid fixation and spread of mutations during early tumorigenesis.

The dynamic nature of methylation patterns severely restricts their use as a clonal marker *per-se*, as demonstrated in Chapter 5. However, in situations when clonality can be inferred by an alternative means, such as by rare somatic point mutations in mtDNA, the additional analysis of methylation patterns of non-expressed genes provide an attractive means to estimate the rates of recent clonal expansions. Thus, the principles and techniques presented here could easily be applied to study the dynamics of clonal expansion in many other human tissues.

All the human intestinal tissue used in Chapters 3, 4 and 5 was obtained form patients who had undergone bowel resections for either colorectal polyps or cancer, and morphological normal tissue was obtained from outside of the resection margins. It has been reported that the colonic epithelium in patients with colorectal cancer demonstrates a generalised increase in cell proliferation [197] and thus, it could be argued that the tissue used in these studies was not 'normal'. However, Potten and colleagues [198-200] have demonstrated variable results when analysing morphological normal crypts both immediately adjacent to and at a distance from tumour sites, initially reporting generally depressed proliferative activity in crypts adjacent to tumours [198, 199] and then observing no evidence of any proliferative differences in a subsequent study [200]. In the study by Becciolini et al [200], morphologically normal colon crypts taken at a distance of both 5 cm and 20cm from above and below tumours demonstrated no differences when assessing crypt total cell number, the ^{3}H thymidine labelling index and mitotic indices, in addition normal colon tissue from cancer patients did not demonstrate the generalised increase in proliferative activity reported by Terpstra et al [197].

Despite colorectal adenomas being one of the most clinically relevant pre-neoplastic lesions in epithelia, and vast amounts of expertise and money currently being invested in establishing widespread endoscopic screening programmes, very little is known about how they grow. Large adenomas, greater than 1cm, have been shown to have a significant malignant risk [153], but the clinical relevance of smaller lesions has been the subject of much debate due to studies suggesting they show little or static growth over many years [155, 158-160]. Furthermore, the follow up of patients with adenomas detected at colonoscopy has been predominantly determined by non-evidence based opinions of best practice. The data from Chapter 6 suggests that there may be rapid clonal expansion occurring within adenomas, both over small areas of adenomatous crypts and involving the entire lesion, and that adenomas grow in a punctuated manner. Therefore, the macroscopic size of an adenoma may be clinically irrelevant, since even small adenomas that appear to be quiescent may possess significant malignant potential. Further work, perhaps incorporating next generation sequencing techniques that enable the rapid

sequencing of large amounts of the genome, could be applied to adenomas in order to better define the dynamics of genetically distinct sub-clones within adenomas, and this may allow improved stratification of risk in patients with polyps detected at endoscopy.

There are a number of limitations to the data presented in this study. The numbers of samples analysed are small, and further work needs to be done to ensure a representative data set has been obtained. The estimates of the dynamics of clonal expansion using the methylation status at CpG islands of non-expressed genes, and the mathematical model employed in Chapter 5, depend on the validity of the various assumptions made, thus the data has to be interpreted with caution. Importantly, observational data has inherent bias associated with study design and data collection, analysis and interpretation, therefore further work to validate the conclusions, perhaps using experimental mouse models or tissue culture methods, is needed.

In summary, more work needs to be done to understand the stem cell biology of the intestine, and how this may be altered in very early tumorigenesis, as this may be crucial in enabling the development of targeted chemoprevention for colorectal cancer, and further studies to delineate the dynamics of adenoma growth should help to better inform the clinical management of patients with these important lesions.

Chapter 9. Appendix

9.1 PCR primer sequences and reaction conditions

All primers were ordered dry from Sigma-Aldrich, UK, and re-suspended in water to form a 20 micromolar working solution.

9.1.1 Primer details for sequencing of entire mitochondrial genome

Name	Sequence 5' to 3'
A(F)	GCTCACATCACCCCATAAAC
A(R)	GATTACTCCGGTCTGAACTC
B(F)	ACCAACAAGTCATTATTACCC
B(R)	TGAGGAAATACTTGATGGCAG
C(F)	CCGTCATCTACTCTACCATC
C(R)	GGACGGATCAGACGAAGAG
D(F)	AATACCCATCATAATCGGAGG
D(R)	GGTGATGAGGAATAGTGTAAG
E(F)	AACCACTTTCACCGCTACAC
E(R)	AGTGAGATGGTAAATGCTAG
F(F)	ACTTCACGTCATTATTGGCTC
F(R)	ATAGGAGGAGAATGGGGGATAG
G(F)	ACCCCCCACTATTAACCTACTG
G(R)	GGTAGAATCCGAGTATGTTGG
H(F)	TATTCGCAGGATTTCTCATTAC
H(R)	AGCTTTGGGTGCTAATGGTG
I(F)	CCCATCCTCCATATATCCAAAC
I(R)	GGTTAGTATAGCTTAGTTAAAC

Table 9.1 Forward (F) and reverse (R) primers for first-round nested mitochondrial PCR.
Primer pairs span the entire human mitochondrial genome. See table 9.16 for PCR reagent and thermal cycler protocol.

Name	Sequence 5’ to 3’
1(F)	TGTAAAACGACGGCCAGTCACACACACCGCTGCTAAC
2(F)	TGTAAAACGACGGCCAGTTTAAAACTCAAAGGACCTGGC
3(F)	TGTAAAACGACGGCCAGTAACTTAACTTGACCGCTCTGAG
4(F)	TGTAAAACGACGGCCAGTACTGTTAGTCCAAAGAGGAAC
5(F)	TGTAAAACGACGGCCAGTCAGTGACACATGTTTAACGGC
6(F)	TGTAAAACGACGGCCAGTCAGCCGCTATTAAAGGTTCG
7(F)	TGTAAAACGACGGCCAGTACCATCACCCTCTACATCAC
8(F)	TGTAAAACGACGGCCAGTTCGCCCTATTCTTCATAGCC
9(F)	TGTAAAACGACGGCCAGTACACTCATCACAGCGCTAAG
10(F)	TGTAAAACGACGGCCAGTCTCACTCTCTCAATCTTATCC
11(F)	TGTAAAACGACGGCCAGTACCTCAATCACACTACTCCC
12(F)	TGTAAAACGACGGCCAGTAGATTTACAGTCCAATGCTTC
13(F)	TGTAAAACGACGGCCAGTTAGCAGGTGTCTCCTCTATC
14(F)	TGTAAAACGACGGCCAGTATTTAGCTGACTCGCCACAC
15(F)	TGTAAAACGACGGCCAGTGGCTCATTCATTTCTCTAACAG
16(F)	TGTAAAACGACGGCCAGTTCCTAACACTCACAACAAAAC
17(F)	TGTAAAACGACGGCCAGTACAGTTTCATGCCCATCGTC
18(F)	TGTAAAACGACGGCCAGTACCACCCAACAATGACTAATC
19(F)	TGTAAAACGACGGCCAGTATCCTAGAAATCGCTGTCGC
20(F)	TGTAAAACGACGGCCAGTCATCCGTATTACTCGCATCAG
21(F)	TGTAAAACGACGGCCAGTCAACACCCTCCTAGCCTTAC
22(F)	TGTAAAACGACGGCCAGTATCGCTCACACCTCATATCC
23(F)	TGTAAAACGACGGCCAGTTATCCAGTGAACCACTATCAC
24(F)	TGTAAAACGACGGCCAGTTCCTTGTACTATCCCTATGAG
25(F)	TGTAAAACGACGGCCAGTCTCCCTCTACATATTTACCAC
26(F)	TGTAAAACGACGGCCAGTCTCTTCCCCACAACAATATTC
27(F)	TGTAAAACGACGGCCAGTGCCCTTCTAAACGCTAATCC
28(F)	TGTAAAACGACGGCCAGTCGGGTCCATCATCCACAAC
29(F)	TGTAAAACGACGGCCAGTACCTAAAACTCACAGCCCTC
30(F)	TGTAAAACGACGGCCAGTATTAAAGTTTACCACAACCACC
31(F)	TGTAAAACGACGGCCAGTATTCATCGACCTCCCCACC
32(F)	TGTAAAACGACGGCCAGTCATCTTGCCCTTCATTATTGC
D1(F)	TGTAAAACGACGGCCAGTATCGGAGGACAACCAGTAAG
D2(F)	TGTAAAACGACGGCCAGTCTCAACTATCACACATCAACTG
D3(F)	TGTAAAACGACGGCCAGTCCTTAAATAAGACATCACGATG
D4(F)	TGTAAAACGACGGCCAGTGCCACAGCACTTAAACACATC

Table 9.2 Forward (F) primers for second-round nested mitochondrial PCR.
Primers are specific for the amplicons from the first-round reaction. See table 9.17 for PCR reagent and thermal cycler protocols.

Name	Sequence 5’ to 3’
1(R)	CAGGAAACAGCTATGACCGATGGCGGTATATAGGCTGAG
2(R)	CAGGAAACAGCTATGACCCTGGTAGTAAGGTGGAGTGGG
3(R)	CAGGAAACAGCTATGACCATTGGTGGCTGCTTTTAGG
4(R)	CAGGAAACAGCTATGACCTCGTGGAGCCATTCATACAG
5(R)	CAGGAAACAGCTATGACCGATTACTCCGGTCTGAACTC
6(R)	CAGGAAACAGCTATGACCGGAGGGGGGTTCATAGTAG
7(R)	CAGGAAACAGCTATGACCAGAGTGCGTCATATGTTGTTC
8(R)	CAGGAAACAGCTATGACCGTTTATTTCTAGGCCTACTCAG
9(R)	CAGGAAACAGCTATGACCGATTTTGCGTAGCTGGGTTTG
10(R)	CAGGAAACAGCTATGACCTGTAGGAGTAGCGTGGTAAGG
11(R)	CAGGAAACAGCTATGACCTAGTCAACGGTCGGCGAAC
12(R)	CAGGAAACAGCTATGACCATGGCAGGGGGTTTTATATTG
13(R)	CAGGAAACAGCTATGACCAAGAAAGATGAATCCTAGGGC
14(R)	CAGGAAACAGCTATGACCCATCCATATAGTCACTCCAGG
15(R)	CAGGAAACAGCTATGACCGGCAGGATAGTTCAGACGG
16(R)	CAGGAAACAGCTATGACCTACAGTGGGCTCTAGAGGG
17(R)	CAGGAAACAGCTATGACCGTATAAGAGATCAGGTTCGTC
18(R)	CAGGAAACAGCTATGACCGTTGTCGTGCAGGTAGAGG
19(R)	CAGGAAACAGCTATGACCATTAGACTATGGTGAGCTCAG
20(R)	CAGGAAACAGCTATGACCTAGCCGTTGAGTTGTGGTAG
21(R)	CAGGAAACAGCTATGACCAGGCACAATATTGGCTAAGAG
22(R)	CAGGAAACAGCTATGACCATGATTAGTTCTGTGGCTGTG
23(R)	CAGGAAACAGCTATGACCTAGGTCTGTTTGTCGTAGGC
24(R)	CAGGAAACAGCTATGACCCGTGTGAATGAGGGTTTTATG
25(R)	CAGGAAACAGCTATGACCGTGGCTCAGTGTCAGTTCG
26(R)	CAGGAAACAGCTATGACCCTGATTTGCCTGCTGCTGC
27(R)	CAGGAAACAGCTATGACCGGGAGGTTGAAGTGAGAGG
28(R)	CAGGAAACAGCTATGACCGTTAGGTAGTTGAGGTCTAGG
29(R)	CAGGAAACAGCTATGACCAGGATTGGTGCTGTGGGTG
30(R)	CAGGAAACAGCTATGACCAAGGAGTGAGCCGAAGTTTC
31(R)	CAGGAAACAGCTATGACCGGTTGTTTGATCCCGTTTCG
32(R)	CAGGAAACAGCTATGACCTACAAGGACAGGCCCATTTG
D1(R)	CAGGAAACAGCTATGACCAGGGTGATAGACCTGTGATC
D2(R)	CAGGAAACAGCTATGACCAGATACTGCGACATAGGGTG
D3(R)	CAGGAAACAGCTATGACCCTGGTTAGGCTGGTGTTAGG
D4(R)	CAGGAAACAGCTATGACCTGCTGCGTGCTTGATGCTTG

Table 9.3 Reverse (R) primers for second-round nested mitochondrial PCR.
Primers are specific for the amplicons from the first-round reaction. See table 9.17 for PCR reagent and thermal cycler protocols.

9.1.2 Primer details and reaction conditions for RT-PCR

Name	Sequence 5' to 3'	Reaction notes
Villin1st F	CAAGACAGGCTCACTCACCA	Mg 2.5/60˚C
Villin 1st R	ATCCCCTCGGTTGAAACTCT	
Villin 2nd F	CACCTTTGGAAGCTTCTTCG	Mg 1.5/60˚C
Villin 2nd R	TCCAGCTACCACGTTCCTCT	
CD-10 1st F	GCTGAGGGGTCACGATTTTA	Mg 2.5/60˚C
CD-10 1st R	AGCTGTCCAAGAAGCACCAT	
CD-10 2nd F	TCCACTGCAGATCAGCCTCT	Mg 2.5Q/55˚C
CD-10 2nd R	TGCTTTTGCTTTCTGCACTG	
Muc-2 1st F	CTCCAGACAGAGGGCAGAAC	Mg 2.5/60˚C
Muc-2 1st R	GATCCTCACACACCACATCG	
Muc-2 2nd F	AAGGAATTTGCTGTGCACCT	Mg 2.5Q/60˚C
Muc-2 2nd R	GCCGTCAGAGAGGAATTCTG	
Chrom A 1st F	GGGATACCGAGGTGATGAAA	Mg 2.5Q/60˚C
Chrom A 1st R	TCCACCAGACCCTGAGAGAG	
Chrom A 2nd F	TTGAGGTCATCTCCGACACA	Mg 2.5/60˚C
Chrom A 2nd R	TTTTCTCTGCCTCCTTGGAA	
β-Actin F	AGCACAGAGCCTCGCCTTT	Mg 2.5Q/60˚C
β-Actin R	CCCACATAGGAATCCTTCTGAC	

Table 9.4 Lineage specific primers for cDNA nested PCR
Primers were designed from mRNA and are all exon-exon spanning and separated by at least one intron. **1st** and **2nd** refer to PCR reaction round; **F:** Forward primer; **R:** Reverse primer. Reaction notes refer to standard PCR reagent and thermal cycler protocols (see Tables 9.18 and Table 9.21).

9.1.3 Primer details and reaction conditions for PCR, cloning and sequencing of target CpG loci within *CSX, BGN* and *MYOD1*

Name	Sequence 5' to 3'	Reaction notes
CSX 1st F	GAGTTTGGTAGGGAAGGGATT	Mg 1.5/55˚C
CSX 1st R	AAAACACTCCTAAAAAAACAACTA	
CSX 2nd F	GGAGATTTAGGAATTTTTTTTGTTTT	Mg 1.5/60˚C
CSX 2nd R	CACCAAACTACAAAATCACTCATTACA	
BGN 1st F	TTTTTTTGAAGTTGTTAGGG	Mg 1.5/55˚C
BGN 1st R	AACCAAAAACATCTCTAAATTACT	
BGN 2nd F	TAAATTGTTTAGGAGTGAGTAGTTGTTT	Mg 2.5/60˚C
BGN 2nd R	AAAAACAACTTAAAACCAACCCTACC	
MYOD 1st F	GGGTTTTTTTTTTAGTTGAAGAGGT	Mg 2.5Q/60˚C
MYOD 1st R	ACCTAAAAATTACTCAACAAAT	
MYOD 2ndF	TGGAGGGGATTTTTAATTTGG	Mg 1.5Q/55˚C
MYOD 2nd R	AACCCAATCCTTCTTCCCTAA	
M13F	GTAAAACGACGGCCAGT	Mg 1.5/55˚C
M13R	CAGGAAACAGCTATGAC	
T7	TAATACGACTCACTATAGGG	Mg 1.5/55˚C
SP6	TATTTAGGTGACACTATAG	

Table 9.5 Methylation nested PCR, cloning and sequencing primers
1st and **2nd** refer to PCR reaction round; primers are specific for bisulphite treated DNA. **F:** Forward primer; **R:** Reverse primer. Reaction notes refer to standard PCR reagent and thermal cycler protocols (see Table 9.18 and Table 9.21).

9.1.4 Primer details and reaction conditions for *APC*, *KRAS*, *BRAF* and *TP53* sequenced loci

Name	Sequence 5' to 3'	Reaction notes
APC MCR1 F	GAGTTTGGTAGGGAAGGGATT	Mg 2.5Q/60˚C
APC MCR1 R	AAAACACTCCTAAAAAAACAACTA	
APC MCR2 F	GGAGATTTAGGAATTTTTTTTGTTTT	Mg 2.5Q/55˚C
APC MCR2R	CACCAAACTACAAAATCACTCATTACA	

Table 9.6 *APC* mutation cluster region (MCR) PCR primer sequences for frozen tissue
MCR1 and MCR2 primers span the *APC* MCR (codons 1250-1450) and produce two amplicons of approximately 600kb. Reaction notes refer to standard PCR reagent and thermal cycler protocols (see Table 9.18 and Table 9.21).

Name	Sequence 5’ to 3’	Reaction notes
APC 1st 1F	GGACAAAGCAGTAAAACCGAAC	Mg 1/55˚C
APC 1st 1R	AACTACATCTTGAAAAACATATTGGA	
APC 1st 2F	AAGTGGTCAGCCTCAAAAGG	Mg 3Q/60˚C
APC 1st 2R	GCTATTTGCAGGGTATTAGCA	
APC 1st 3F	GATACTCCAATATGTTTTTCAAGATG	Mg 1/55˚C
APC 1st 3R	GCCTGGCTGATTCTGAAGAT	
APC 1st 4F	CCCTGCAAATAGCAGAAATAAAA	Mg 1/55˚C
APC 1st 4R	AACATGAGTGGGGTCTCCTG	
APC 1st 5F	CAGACTGCAGGGTTCTAGTTTATC	Mg 2/60˚C
APC 1st 5R	CATTCCACTGCATGGTTCAC	
APC 1st 6F	CCAAAAGTGGTGCTCAGACA	Mg 2/60˚C
APC 1st 6R	CATGGTTTGTCCAGGGCTAT	
APC 1st 7F	TTTGAGAGTCGTTCGATTGC	Mg 1/60˚C
APC 1st 7R	TCTCTTTTCAGCAGTAGGTGCTT	
APC 1st 8F	CATGCAGTGGAATGGTAAGT	Mg 2/55˚C
APC 1st 8R	GCAGCATTTACTGCAGCTT	
APC 1st 9F	CAAGCGAGAAGTACCTAAAAA	Mg 2/55˚C
APC 1st 9R	TTCTGTATAAATGGCTCATCG	
APC 1st 10F	GGTTCTTCCAGATGCTGATA	Mg 2/55˚C
APC 1st 10R	CTTGGTTTTCATTTGATTCTTT	
APC 1st 11F	AATTAAGAATAATGCCTCCAGT	Mg 2/55˚C
APC 1st 11R	TTTACGTGATGACTTTGTTGG	
APC 1st 12F	CCAAGAGAAAGAGGCAGAAAAA	Mg 1Q/55˚C
APC 1st 12R	TGATGGTAGAAGTTTGTACACAGG	

Table 9.7 *APC* MCR PCR nested first-round primer sequences
Primers span the *APC* MCR (codons 1250-1450). Reaction notes refer to standard PCR reagent and thermal cycler protocols (see Table 9.18 and Table 9.21).

Name	Sequence 5’ to 3’	Reaction notes
APC 2nd 1F	CAAGCAGTGAGAATACGTCCA	Mg 1/55˚C
APC 2nd 1R	TTTCTTGGTTAATAGAAGAAACTTTGC	
APC 2nd 2F	GCCACTTGCAAAGTTTCTTCT	Mg 2Q/60˚C
APC 2nd 2R	TGCTTCCTGTGTCGTCTGA	
APC 2nd 3F	CAGACGACACAGGAAGCAGA	Mg 1 /55˚C
APC 2nd 3R	TGGAACTTCGCTCACAGGAT	
APC 2nd 4F	GAAGATCCTGTGAGCGAAGTTCC	Mg 1/55˚C
APC 2nd 4R	CAAGTCCTCTGGGGTGAGTA	
APC 2nd 5F	AGAATCAGCCAGGCACAAAG	Mg 1/60˚C
APC 2nd 5R	GCAATCGAACGACTCTCAAA	
APC 2nd 6F	CACTATGTTCAGGAGACCCCA	Mg 1/60˚C
APC 2nd 6R	TGGAAGATCACTGGGGCTTA	
APC 2nd 7F	GTGAACCATGCAGTGGAATG	Mg 1/60˚C
APC 2nd 7R	ACTTCTCGCTTGGTTTGAGC	
APC 2nd 8F	GCTTAGGTCCACTCTCTCTCTT	Mg 1/55˚C
APC 2nd 8R	GGCATTATAAGCCCCAGT	
APC 2nd 9F	CCTAAAAATAAAGCACCTACTGCTG	Mg 3/55˚C
APC 2nd 9R	CACTCAGGCTGGATGAACAA	
APC 2nd 10F	ACATTTTGCCACGGAAAGTA	Mg 2/55˚C
APC 2nd 10R	GGCTGCTCTGATTCTGTTTC	
APC 2nd 11F	CAGGAAAATGACAATGGGAAT	Mg 2/55˚C
APC 2nd 11R	TGGCATGGCAGAAATAATACA	
APC 2nd 12F	GGACCTATTAGATGATTCAGATGATG	Mg 2Q/55˚C
APC 2nd 12R	ACTTGGTTTCCTTGCCACAG	

Table 9.8 *APC* MCR PCR nested second-round primer sequences and reaction notes
Primers are specific for the first-round amplicons. Reaction notes refer to standard PCR reagent and thermal cycler protocols (see Table 9.18 and Table 9.21).

Name	Sequence 5' to 3'	Reaction notes
KRAS 1 F	GAGTTTGTATTAAAAGGTACTGGTGGA	Mg 2Q/60˚C
KRAS 1 R	ATCAAAGAATGGTCCTGCAC	
KRAS 2 F	TTTGATAGTGTATTAACCTTAT	Mg 2Q/55˚C
KRAS 2R	TATTAAAACAAGATTTACCTC	

Table 9.9 *KRAS* nested PCR primer sequences
1st and **2nd** refer to PCR reaction round; **F:** Forward primer; **R:** Reverse primer. Reaction notes refer to specific PCR reagent and thermal cycler protocols detailed in Table 9.19 and Table 9.21.

Name	Sequence 5' to 3'	Reaction notes
BRAF F	TCATAATGCTTGCTCTGATAGGA	Mg 1.5 /55˚C
BRAF R	GGCCAAAAATTTAATCAGTGGA	

Table 9.10 *BRAF* PCR primer sequences
F: Forward primer; **R:** Reverse primer. Reaction notes refer to standard PCR reagent and thermal cycler protocols (see Table 9.18 and Table 9.21).

Name	Sequence 5’ to 3’	Reaction notes
TP53 5 1F	CACTTGTGCCCTGACTTTCA	Mg 1Q /55˚C
TP53 5 1R	GAGCAATCAGTGAGGAATCAGA	
TP53 5 2F	TCTGTCTCCTTCCTCTTCCTACA	Mg 1Q /60˚C
TP53 5 2R	AACCAGCCCTGTCGTCTCT	
TP53 6 1F	AGAGACGACAGGGCTGGTT	Mg 2Q /60˚C
TP53 6 1R	TGGAGGGCCACTGACAAC	
TP53 6 2F	CAGGCCTCTGATTCCTCACT	Mg 1/60˚C
TP53 6 2R	CTTAACCCCTCCTCCCAGAG	
TP53 7 1F	TGCTTGCCACAGGTCTCC	Mg 1Q /60˚C
TP53 7 1R	GGTCAGAGGCAAGCAGAGG	
TP53 7 2F	CTTGGGCCTGTGTTATCTCC	Mg 1Q /60˚C
TP53 7 2R	GTGTGCAGGGTGGCAAGT	
TP53 8 1F	TTTTTAAATGGGACAGGTAGGA	Mg 2Q /60˚C
TP53 8 1R	CACCCTTGGTCTCCTCCAC	
TP53 8 2F	GCCTCTTGCTTCTCTTTTCC	Mg 2/60˚C
TP53 8 2R	GCTTCTTGTCCTGCTTGCTT	

Table 9.11 *TP53* nested PCR primer sequences
1st and **2nd** refer to PCR reaction round; **F:** Forward primer; **R:** Reverse primer. Reaction notes refer to specific PCR reagent and thermal cycler protocols detailed in Tables 9.19 and Table 9.21.

9.1.5 Microsatellite loss of heterozygosity analysis: primers and reaction details for multiplex PCR

Primer	Sequence 5' to 3'	Reaction notes
Multiplex 1		
D5S2001 F	GCCAAGATGGTCTCGATCTC	MP57
D5S2001 R	TCTGAACAGGTGATGGCAAC	
D17S1832 F	ACGCCTTGACATAGTTGC	
D17S1832 R	TGTGTGACTGTTCAGCCTC	
D5S346 F	ACTCACTCTAGTGATAAATCGGG	
D5S346 R	AGCAGATAAGACAGTATTACTAGTT	
Multiplex 2		
D5S489 F	GGGCTTTTGTGTTGTTTCTA	MP57
D5S489 R	GAAAACCCATAACCAGACTTG	

Table 9.12 Microsatellite LOH primers
F: Forward primer; **R:** Reverse primer. Reaction notes refer to multiplex PCR reagent and thermal cycler protocol (MP57) detailed in Table 9.25 and annealing temperatures.

9.1.6 Microsatellite instability (MSI) primers and reaction details

Name	Sequence 5' to 3'	Reaction notes
BAT25 F	[HEX]TCGCCTCCAAGAATGTAAGT	MP57˚C
BAT25 R	TCTGCATTTTAACTATGGCTC	
BAT26 F	[6FAM]TGACTACTTTTGACTTCAGAA	MP57˚C
BAT26 R	AACCATTCAACATTTTTAACCC	

Table 9.13 MSI primers for multiplex PCR
F: Forward primer; **R:** Reverse primer. Reaction notes refer to specific multiplex PCR reagent and thermal cycler protocol (MP57) detailed in Table 9.25 and annealing temperatures.

9.2 PCR reagent and thermal cycling protocols

All volumes are in microlitres (μl) and temperatures in °C unless otherwise stated.

Reagent	Manufacturer	Working concentration
Nucleotides	Promega, UK	10mM of each
Buffer	Promega, UK	As supplied
Magnesium	Qiagen, UK	25mM
Taq	Qiagen, UK	5units/μl
Q solution	Qiagen, UK	5x (as supplied)

Table 9.14 Standard PCR reagents
The above reagents were used in all PCR reactions other than the mitochondrial DNA PCR.

Reagent	Manufacturer	Working concentration
Nucleotides	Applied Biosystems	10mM of each
Buffer*	Applied Biosystems	As supplied
Magnesium	Applied Biosystems	25mM
AmpliTaq Gold	Applied Biosystems	5units/μl

Table 9.15 MtDNA PCR reagents
The above reagents were used in all mtDNA PCR reactions. *Buffer contained 15mM magnesium.

First-round mtDNA PCR protocol				
Reagents per well (total = 50 μl)		Thermal cycler programme		
H_20	33.65 μl	**95°C**	10 min	
Nucleotides	5.0 μl	**94°C**	45 sec	38 cycles
Buffer	5.0 μl	**58°C**	45 sec	
Mg^{2+} (25mM)	2.0 μl	**72°C**	2 min	
AmpliTaq gold	0.35 μl	**72°C**	8 min	
DNA lysate	1.0 μl	**4°C**	thereafter	
Primers	1.5 μl of each			

Table 9.16 Mitochondrial DNA PCR first-round protocols
Reagent and thermal cycler protocols detailed.

Second-round mtDNA PCR protocol				
Reagents per well (total = 25 μl)		Thermal cycler programme		
H_20	16.87 μl	**95°C**	10 min	
Nucleotides	2.5 μl	**94°C**	45 sec	30 cycles
Buffer	2.5 μl	**58°C**	45 sec	
AmpliTaq gold	0.35 μl	**72°C**	1 min	
DNA lysate	1.0 μl	**72°C**	8 min	
Primers	1.0 μl of each	**4°C**	thereafter	

Table 9.17 Mitochondrial DNA PCR second-round protocols
Reagent and thermal cycler protocols detailed.

Standard PCR reagent protocols										
Reagents per well **(all volumes in µl, total = 25 µl per well)**										
	Mg 1	**Mg1.5**	**Mg 2**	**Mg2.5**	**Mg3**	**Mg1 Q**	**Mg1.5Q**	**Mg2 Q**	**Mg2.5Q**	**Mg3 Q**
H_20	18.25	17.75	17.25	16.75	16.25	13.25	12.75	12.25	11.75	11.25
Nucleotides	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5
Buffer	2.5	2.5	2.5	2.5	2.5	2.5	2.5	2.5	2.5	2.5
Mg^{2+} (25mM)	1.0	1.5	2.0	2.5	3.0	1.0	1.5	2.0	2.5	3.0
Q	0	0	0	0	0	5	5	5	5	5
Taq	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25
DNA lysate	2.0	2.0	2.0	2.0	2.0	2.0	2.0	2.0	2.0	2.0
(F) primer	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25
(R) primer	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25

Table 9.18 Standard PCR reagent protocols
PCR reaction conditions contained either 1.0mM, 1.5mM, 2.0mM, 2.5mM or 3.0mM magnesium concentration, with or without Q solution (Q). Thermal cycler programmes were run with either a 55°C or 60°C annealing temperature depending on the primers used. See Appendix 9.1 for reaction notes specific to each primer set. (**F):** Forward primer; (**R):** Reverse primer.

***KRAS/TP53* PCR reagent protocols**				
Reagents per well (all volumes in µl, total = 25 µl per well)				
	Mg 1	**Mg2**	**Mg1Q**	**Mg2Q**
H_20	18.3	17.3	13.3	12.3
Nucleotides	0.5	0.5	0.5	0.5
Buffer	2.5	2.5	2.5	2.5
Mg^{2+} (25mM)	1.0	2.0	1.0	2.0
Q	0	0	5	5
Taq	0.2	0.2	0.2	0.2
DNA lysate	2.0	2.0	2.0	2.0
(F) primer	0.25	0.25	0.25	0.25
(R) primer	0.25	0.25	0.25	0.25

Table 9.19 *KRAS*/*TP53* PCR reagent protocols
PCR reaction conditions contained either 1.0mM or 2.0mM magnesium concentration, with or without Q solution (Q). Thermal cycler programmes were run with either a 55°C or 60°C annealing temperature depending on the primers used. See Appendix 9.1 for reaction notes specific to each primer set. (**F):** Forward primer; (**R):** Reverse primer.

DNA sequencing reaction reagent protocol	
Reagents per well (all volumes in μl, total = 20 μl per well)	
BDT	10
H_20	5
Primer*	1
PCR product**	4

Table 9.20 PCR product sequencing reaction protocol
BDT: BigDye Terminator (Applied Biosystems, USA); *forward or reverse primer used; **diluted ExoSap-IT purified PCR product.

Standard thermal cycler PCR programme			
Step	**Temperature**	**Time**	
Temperature 1	95°C	5 min	
Denaturing	94°C	1 min	35 cycles
Annealing*	55, 57 or 60°C*	1 min	
Elongation	72°C	1 min	
Temperature 2	72°C	10 min	
Hold	4°C	thereafter	

Table 9.21 Standard thermal cycler programmes
This programme was used for all PCR reactions other than the mtDNA PCR. *See Appendix 9.1 for annealing temperatures specific to each primer set.

ExoSap-IT thermal cycler programme		
Step	**Temperature**	**Time**
Temperature 1	37°C	15 min
Temperature 2	80°C	15 min
Hold	15°C	thereafter

Table 9.22 ExoSap thermal cycler programme

Standard thermal cycler DNA sequencing programme			
Step	**Temperature**	**Time**	
Temperature 1	96°C	5 min	
Denaturing	96°C	30 sec	35 cycles
Annealing	50°C	45 sec	
Elongation	60°C	4 min	
Hold	8°C	thereafter	

Table 9.23 Thermal cycling programme for standard sequencing of PCR products
Regents were prepared as detailed in table 9.20.

mtDNA thermal cycler sequencing programme			
Step	**Temperature**	**Time**	
Temperature 1	96°C	1 min	
Denaturing	96°C	10 sec	25 cycles
Annealing	50°C	5 sec	
Elongation	60°C	4 min	
Hold	4°C	thereafter	

Table 9.24 Thermal cycling programme for sequencing of mtDNA PCR products
Regents were prepared as detailed in table 9.20.

Multiplex PCR protocol (MP57)				
Reagents per well (total = 50 μl)		Thermal cycler programme		
H_20	18μ	**95°C**	10 min	
Primer mix	5μl	**95°C**	45 sec	30 cycles
Multiplex mastermix	25μ	**57°C**	45 sec	
DNA template	2μ	**95°C**	1 min	
		72°C	8 min	
		4°C	thereafter	

Table 9.25 Multiplex PCR protocol MP57
A commercial kit multiplex kit (Qiagen, UK) was used.

9.3 Methylation gene CpG island target sequences

Locus	Sequence
BGN (10 CpGs)	TaaaTtgTTTaggagtgagtagTtgTttt**CG**gtT**CG**T**CG**gaTaTaT**CG**gaTagatag a**CG**tg**CG**ga**CG**gTTTaTTaTTTTagTT**CG**TTaaTtagtTagTTtg**CG**TTtgg **CG**TTtTTTTtTtTTaggtagggTtggTttTaagTtgTTtTT
CSX (8 CpGs)	**G**gagaTTtaggaaTtttttTtgtTTTa**CGCGCG**tttgttTttg**CG**Ta**CG**ggagagtttgt ggg**CG**g**CG**attatgTag**CG**tgTaatgagtgatTTtgTagTTtggtg
MYOD (5 CpGs)	tggaggggattTTtaaTTtgggTaggatT**CG**agtttggagagattgg**CGCG**aagTtttagT agTaatTtT**CG**attTTtgtaTaaTTatagTtgggtttTtaag**CG**tTtagggaagaaggaTtg ggTT

Table 9.26 Reference sequences of *CSX*, *BGN* and *MYOD*
Sequences are displayed post-bisulphite treatment. A capital T indicates a thymine base produced by bisulphite conversion of the original cytosine base. CpG sites are assumed methylated and indicated in **bold**. Second round primer binding sites are underlined.

9.4 Patient details

ID	Age	Sex	Reason for resection
1	78	M	Colorectal cancer
2	73	M	Tubulo-villous adenoma
3	75	F	Colorectal cancer
4	76	F	Colorectal cancer
5	68	M	Ileo-caecal valve polyp
6	73	M	Tubulo-villous adenoma
7	54	F	Colorectal cancer
8	76	F	Colorectal cancer
9	83	M	Colorectal cancer
10	73	M	Polyp
11	61	M	Polyp
12	66	M	Polyp
13	64	M	Polyp
14	69	F	Polyp
15	62	F	Polyp
16	60	M	Polyp
17	65	F	Polyp
18	64	M	Polyp
19	60	M	Polyp

Table 9.27 Patient details for normal human colonic tissue and resected polyps from Chapters 5 and 6
For patients 1-9 fresh frozen normal mucosa was collected at least 15cm from the cancer margin. Fresh frozen polyps and archived FFPE tissue was collected for patients 10-19.

9.5 Publications

Work presented within this thesis has been published in the following, international, peer-reviewed scientific journals:

- Humphries A, Wright NA. Colonic crypt organisation and tumorigenesis. Nature Rev Cancer. 2008 Jun;8(6):415-24. Epub 2008 May 15. Review.

- Fellous TG, McDonald SA, Burkert J, Humphries A, Islam S, De-Alwis NM, Gutierrez-Gonzalez L, Tadrous PJ, Elia G, Kocher HM, Bhattacharya S, Mears L, El-Bahrawy M, Turnbull DM, Taylor RW, Greaves LC, Chinnery PF, Day CP, Wright NA, Alison MR. A methodological approach to tracing cell lineage in human epithelial tissues. Stem Cells. 2009 Mar 19;27(6):1410-1420. [Epub ahead of print].

- Graham TA, Humphries A, Sanders T, Rodriguez-Justo M, Tadrous PJ, Preston SL, Novelli MR, Leedham SJ, McDonald SA, Wright NA . Use of Methylation Patterns to Determine Expansion of Stem Cell Clones in Human Colon Tissue. Gastroenterology 2011 Apr;140(4):1241-1250.e9. Epub 2010 Dec 28.

References

1. Ferlay, J. et al. Estimates of worldwide burden of cancer in 2008: Globocan 2008. *Int J Cancer* (2010).
2. Statistics, O.f.N. Cancer Statistics registrations: registrations of cancer diagnosed in 2008, England. (2011).
3. Salzman, N.H., Underwood, M.A. & Bevins, C.L. Paneth cells, defensins, and the commensal microbiota: a hypothesis on intimate interplay at the intestinal mucosa. *Semin Immunol* **19**, 70-83 (2007).
4. Sato, T. et al. Paneth cells constitute the niche for Lgr5 stem cells in intestinal crypts. *Nature* **469**, 415-8 (2011).
5. Cheng, H. & Leblond, C.P. Origin, differentiation and renewal of the four main epithelial cell types in the mouse small intestine. V. Unitarian Theory of the origin of the four epithelial cell types. *Am J Anat* **141**, 537-61 (1974).
6. Ponder, B.A. et al. Derivation of mouse intestinal crypts from single progenitor cells. *Nature* **313**, 689-91 (1985).
7. Thompson, M. et al. Gastric endocrine cells share a clonal origin with other gut cell lineages. *Development* **110**, 477-81 (1990).
8. Crosnier, C., Stamataki, D. & Lewis, J. Organizing cell renewal in the intestine: stem cells, signals and combinatorial control. *Nat Rev Genet* **7**, 349-59 (2006).
9. Wright NA, A.M. The biology of epithelial cell populations. (Oxford University Press, 1984).
10. Williams, E.D., Lowes, A.P., Williams, D. & Williams, G.T. A stem cell niche theory of intestinal crypt maintenance based on a study of somatic mutation in colonic mucosa. *Am J Pathol* **141**, 773-6 (1992).
11. van Es, J.H. et al. Notch/gamma-secretase inhibition turns proliferative cells in intestinal crypts and adenomas into goblet cells. *Nature* **435**, 959-63 (2005).
12. Fevr, T., Robine, S., Louvard, D. & Huelsken, J. Wnt/beta-catenin is essential for intestinal homeostasis and maintenance of intestinal stem cells. *Mol Cell Biol* **27**, 7551-9 (2007).
13. Kosinski, C. et al. Gene expression patterns of human colon tops and basal crypts and BMP antagonists as intestinal stem cell niche factors. *Proc Natl Acad Sci U S A* **104**, 15418-23 (2007).
14. Batlle, E. et al. Beta-catenin and TCF mediate cell positioning in the intestinal epithelium by controlling the expression of EphB/ephrinB. *Cell* **111**, 251-63 (2002).
15. Sato, T., Vries, R.G., Snippert, H.J., van de Wetering, M., Barker, N., Stange, D.E., van Es, J.H., Abo, A., Kujala, P., Peters, P.J., et al. Single Lgr5 stem cells build crypt-villus structures in vitro without a mesenchymal niche. *Nature* (2009).
16. Winton, D.J. & Ponder, B.A. Stem-cell organization in mouse small intestine. *Proc R Soc Lond B Biol Sci* **241**, 13-8 (1990).
17. Bjerknes, M. & Cheng, H. Clonal analysis of mouse intestinal epithelial progenitors. *Gastroenterology* **116**, 7-14 (1999).
18. Yatabe, Y., Tavare, S. & Shibata, D. Investigating stem cells in human colon by using methylation patterns. *Proc Natl Acad Sci U S A* **98**, 10839-44 (2001).

19. Barker, N. et al. Identification of stem cells in small intestine and colon by marker gene Lgr5. *Nature* **449**, 1003-7 (2007).
20. Sangiorgi, E. & Capecchi, M.R. Bmi1 lineage tracing identifies a self-renewing pancreatic acinar cell subpopulation capable of maintaining pancreatic organ homeostasis. *Proc Natl Acad Sci U S A* **106**, 7101-6 (2009).
21. Potten, C.S. et al. Identification of a putative intestinal stem cell and early lineage marker; musashi-1. *Differentiation* **71**, 28-41 (2003).
22. Dekaney, C.M., Rodriguez, J.M., Graul, M.C. & Henning, S.J. Isolation and characterization of a putative intestinal stem cell fraction from mouse jejunum. *Gastroenterology* **129**, 1567-80 (2005).
23. May, R. et al. Identification of a Novel Putative Gastrointestinal Stem Cell and Adenoma Stem Cell Marker: DCAMKL-1 Following Radiation Injury and in APC/Min Mice. *Stem Cells* (2007).
24. Cheng, H. & Leblond, C.P. Origin, differentiation and renewal of the four main epithelial cell types in the mouse small intestine. I. Columnar cell. *Am J Anat* **141**, 461-79 (1974).
25. Bjerknes, M. & Cheng, H. The stem-cell zone of the small intestinal epithelium. I. Evidence from Paneth cells in the adult mouse. *Am J Anat* **160**, 51-63 (1981).
26. Potten, C.S., Kovacs, L. & Hamilton, E. Continuous labelling studies on mouse skin and intestine. *Cell Tissue Kinet* **7**, 271-83 (1974).
27. Potten, C.S., Booth, C. & Pritchard, D.M. The intestinal epithelial stem cell: the mucosal governor. *Int J Exp Pathol* **78**, 219-43 (1997).
28. van der Flier, L.G. et al. Transcription factor achaete scute-like 2 controls intestinal stem cell fate. *Cell* **136**, 903-12 (2009).
29. Montgomery, R.K. et al. Mouse telomerase reverse transcriptase (mTert) expression marks slowly cycling intestinal stem cells. *Proc Natl Acad Sci U S A* **108**, 179-84 (2011).
30. Breault, D.T. et al. Generation of mTert-GFP mice as a model to identify and study tissue progenitor cells. *Proc Natl Acad Sci U S A* **105**, 10420-5 (2008).
31. Li, L. & Clevers, H. Coexistence of quiescent and active adult stem cells in mammals. *Science* **327**, 542-5 (2010).
32. Cai, W.B., Roberts, S.A. & Potten, C.S. The number of clonogenic cells in crypts in three regions of murine large intestine. *Int J Radiat Biol* **71**, 573-9 (1997).
33. Potten, C.S. & Grant, H.K. The relationship between ionizing radiation-induced apoptosis and stem cells in the small and large intestine. *Br J Cancer* **78**, 993-1003 (1998).
34. Spradling, A., Drummond-Barbosa, D. & Kai, T. Stem cells find their niche. *Nature* **414**, 98-104 (2001).
35. Ohlstein, B., Kai, T., Decotto, E. & Spradling, A. The stem cell niche: theme and variations. *Curr Opin Cell Biol* **16**, 693-9 (2004).
36. Li, L. & Xie, T. Stem cell niche: structure and function. *Annu Rev Cell Dev Biol* **21**, 605-31 (2005).
37. Simons, B.D. & Clevers, H. Strategies for homeostatic stem cell self-renewal in adult tissues. *Cell* **145**, 851-62 (2011).
38. Yamashita, Y.M., Fuller, M.T. & Jones, D.L. Signaling in stem cell niches: lessons from the Drosophila germline. *J Cell Sci* **118**, 665-72 (2005).

39. Yamashita, Y.M., Jones, D.L. & Fuller, M.T. Orientation of asymmetric stem cell division by the APC tumor suppressor and centrosome. *Science* **301**, 1547-50 (2003).
40. Watt, F.M. & Hogan, B.L. Out of Eden: stem cells and their niches. *Science* **287**, 1427-30 (2000).
41. Cairns, J. Somatic stem cells and the kinetics of mutagenesis and carcinogenesis. *Proc Natl Acad Sci U S A* **99**, 10567-70 (2002).
42. Lopez-Garcia, C., Klein, A.M., Simons, B.D. & Winton, D.J. Intestinal stem cell replacement follows a pattern of neutral drift. *Science* **330**, 822-5 (2010).
43. Snippert, H.J. et al. Intestinal crypt homeostasis results from neutral competition between symmetrically dividing Lgr5 stem cells. *Cell* **143**, 134-44 (2010).
44. Kim, K.M. & Shibata, D. Methylation reveals a niche: stem cell succession in human colon crypts. *Oncogene* **21**, 5441-9 (2002).
45. Nicolas, P., Kim, K.M., Shibata, D. & Tavare, S. The stem cell population of the human colon crypt: analysis via methylation patterns. *PLoS Comput Biol* **3**, e28 (2007).
46. Ahuja, N., Li, Q., Mohan, A.L., Baylin, S.B. & Issa, J.P. Aging and DNA methylation in colorectal mucosa and cancer. *Cancer Res* **58**, 5489-94 (1998).
47. Issa, J.P. CpG-island methylation in aging and cancer. *Curr Top Microbiol Immunol* **249**, 101-18 (2000).
48. Pfeifer, G.P., Steigerwald, S.D., Hansen, R.S., Gartler, S.M. & Riggs, A.D. Polymerase chain reaction-aided genomic sequencing of an X chromosome-linked CpG island: methylation patterns suggest clonal inheritance, CpG site autonomy, and an explanation of activity state stability. *Proc Natl Acad Sci U S A* **87**, 8252-6 (1990).
49. Silva, A.J., Ward, K. & White, R. Mosaic methylation in clonal tissue. *Dev Biol* **156**, 391-8 (1993).
50. Maley, C.C. et al. Selectively advantageous mutations and hitchhikers in neoplasms: p16 lesions are selected in Barrett's esophagus. *Cancer Res* **64**, 3414-27 (2004).
51. Meineke, F.A., Potten, C.S. & Loeffler, M. Cell migration and organization in the intestinal crypt using a lattice-free model. *Cell Prolif* **34**, 253-66 (2001).
52. Loeffler, M., Birke, A., Winton, D. & Potten, C. Somatic mutation, monoclonality and stochastic models of stem cell organization in the intestinal crypt. *J Theor Biol* **160**, 471-91 (1993).
53. Bhanot, P. et al. A new member of the frizzled family from Drosophila functions as a Wingless receptor. *Nature* **382**, 225-30 (1996).
54. Pinson, K.I., Brennan, J., Monkley, S., Avery, B.J. & Skarnes, W.C. An LDL-receptor-related protein mediates Wnt signalling in mice. *Nature* **407**, 535-8 (2000).
55. Tamai, K. et al. LDL-receptor-related proteins in Wnt signal transduction. *Nature* **407**, 530-5 (2000).
56. Gregorieff, A. et al. Expression pattern of Wnt signaling components in the adult intestine. *Gastroenterology* **129**, 626-38 (2005).
57. Scoville, D.H., Sato, T., He, X.C. & Li, L. Current view: intestinal stem cells and signaling. *Gastroenterology* **134**, 849-64 (2008).
58. Rubinfeld, B. et al. Binding of GSK3beta to the APC-beta-catenin complex and regulation of complex assembly. *Science* **272**, 1023-6 (1996).

59. Amit, S. et al. Axin-mediated CKI phosphorylation of beta-catenin at Ser 45: a molecular switch for the Wnt pathway. *Genes Dev* **16**, 1066-76 (2002).
60. Haramis, A.P. et al. De novo crypt formation and juvenile polyposis on BMP inhibition in mouse intestine. *Science* **303**, 1684-6 (2004).
61. He, X.C. et al. BMP signaling inhibits intestinal stem cell self-renewal through suppression of Wnt-beta-catenin signaling. *Nat Genet* **36**, 1117-21 (2004).
62. Tian, Q., He, X.C., Hood, L. & Li, L. Bridging the BMP and Wnt pathways by PI3 kinase/Akt and 14-3-3zeta. *Cell Cycle* **4**, 215-6 (2005).
63. Korinek, V. et al. Depletion of epithelial stem-cell compartments in the small intestine of mice lacking Tcf-4. *Nat Genet* **19**, 379-83 (1998).
64. Morin, P.J. et al. Activation of beta-catenin-Tcf signaling in colon cancer by mutations in beta-catenin or APC. *Science* **275**, 1787-90 (1997).
65. Pinto, D., Gregorieff, A., Begthel, H. & Clevers, H. Canonical Wnt signals are essential for homeostasis of the intestinal epithelium. *Genes Dev* **17**, 1709-13 (2003).
66. Andreu, P. et al. Crypt-restricted proliferation and commitment to the Paneth cell lineage following Apc loss in the mouse intestine. *Development* **132**, 1443-51 (2005).
67. Akazawa, C., Ishibashi, M., Shimizu, C., Nakanishi, S. & Kageyama, R. A mammalian helix-loop-helix factor structurally related to the product of Drosophila proneural gene atonal is a positive transcriptional regulator expressed in the developing nervous system. *J Biol Chem* **270**, 8730-8 (1995).
68. Yang, Q., Bermingham, N.A., Finegold, M.J. & Zoghbi, H.Y. Requirement of Math1 for secretory cell lineage commitment in the mouse intestine. *Science* **294**, 2155-8 (2001).
69. Jensen, J. et al. Control of endodermal endocrine development by Hes-1. *Nat Genet* **24**, 36-44 (2000).
70. Schmidt, G.H., Winton, D.J. & Ponder, B.A. Development of the pattern of cell renewal in the crypt-villus unit of chimaeric mouse small intestine. *Development* **103**, 785-90 (1988).
71. Park, H.S., Goodlad, R.A. & Wright, N.A. Crypt fission in the small intestine and colon. A mechanism for the emergence of G6PD locus-mutated crypts after treatment with mutagens. *Am J Pathol* **147**, 1416-27 (1995).
72. Kuraguchi, M., Thomas, G.A. & Williams, E.D. Somatic mutation of the glucose-6-phosphate dehydrogenase (g6pd) gene in colonic stem cells and crypt restricted loss of G6PD activity. *Mutat Res* **379**, 69-75 (1997).
73. Novelli, M.R. et al. Polyclonal origin of colonic adenomas in an XO/XY patient with FAP. *Science* **272**, 1187-90 (1996).
74. Campbell, F. et al. Post-irradiation somatic mutation and clonal stabilisation time in the human colon. *Gut* **39**, 569-73 (1996).
75. Novelli, M. et al. X-inactivation patch size in human female tissue confounds the assessment of tumor clonality. *Proc Natl Acad Sci U S A* **100**, 3311-4 (2003).
76. Taylor, R.W. et al. Mitochondrial DNA mutations in human colonic crypt stem cells. *J Clin Invest* **112**, 1351-60 (2003).
77. Greaves, L.C. et al. Mitochondrial DNA mutations are established in human colonic stem cells, and mutated clones expand by crypt fission. *Proc Natl Acad Sci U S A* **103**, 714-9 (2006).

78. McDonald, S.A. et al. Mechanisms of field cancerization in the human stomach: the expansion and spread of mutated gastric stem cells. *Gastroenterology* **134**, 500-10 (2008).
79. Gutierrez-Gonzalez, L. et al. Analysis of the clonal architecture of the human small intestinal epithelium establishes a common stem cell for all lineages and reveals a mechanism for the fixation and spread of mutations. *J Pathol* **217**, 489-96 (2009).
80. Lightowlers, R.N., Chinnery, P.F., Turnbull, D.M. & Howell, N. Mammalian mitochondrial genetics: heredity, heteroplasmy and disease. *Trends Genet* **13**, 450-5 (1997).
81. Sciacco, M., Bonilla, E., Schon, E.A., DiMauro, S. & Moraes, C.T. Distribution of wild-type and common deletion forms of mtDNA in normal and respiration-deficient muscle fibers from patients with mitochondrial myopathy. *Hum Mol Genet* **3**, 13-9 (1994).
82. Nooteboom, M. et al. Age-associated mitochondrial DNA mutations lead to small but significant changes in cell proliferation and apoptosis in human colonic crypts. *Aging Cell* **9**, 96-9 (2010).
83. Wasan, H.S. et al. APC in the regulation of intestinal crypt fission. *J Pathol* **185**, 246-55 (1998).
84. Miyoshi, Y. et al. Somatic mutations of the APC gene in colorectal tumors: mutation cluster region in the APC gene. *Hum Mol Genet* **1**, 229-33 (1992).
85. Boman, B.M. et al. Colonic crypt changes during adenoma development in familial adenomatous polyposis: immunohistochemical evidence for expansion of the crypt base cell population. *Am J Pathol* **165**, 1489-98 (2004).
86. Lamlum, H. et al. APC mutations are sufficient for the growth of early colorectal adenomas. *Proc Natl Acad Sci U S A* **97**, 2225-8 (2000).
87. Nakamura, S. & Kino, I. Morphogenesis of minute adenomas in familial polyposis coli. *J Natl Cancer Inst* **73**, 41-9 (1984).
88. Woda, B.A., Forde, K. & Lane, N. A unicryptal colonic adenoma, the smallest colonic neoplasm yet observed in a non-polyposis individual. *Am J Clin Pathol* **68**, 631-2 (1977).
89. Slaughter, D.P., Southwick, H.W. & Smejkal, W. Field cancerization in oral stratified squamous epithelium; clinical implications of multicentric origin. *Cancer* **6**, 963-8 (1953).
90. Braakhuis, B.J., Tabor, M.P., Kummer, J.A., Leemans, C.R. & Brakenhoff, R.H. A genetic explanation of Slaughter's concept of field cancerization: evidence and clinical implications. *Cancer Res* **63**, 1727-30 (2003).
91. Dakubo, G.D., Jakupciak, J.P., Birch-Machin, M.A. & Parr, R.L. Clinical implications and utility of field cancerization. *Cancer Cell Int* **7**, 2 (2007).
92. Shih, I.M. et al. Top-down morphogenesis of colorectal tumors. *Proc Natl Acad Sci U S A* **98**, 2640-5 (2001).
93. Preston, S.L. et al. Bottom-up histogenesis of colorectal adenomas: origin in the monocryptal adenoma and initial expansion by crypt fission. *Cancer Res* **63**, 3819-25 (2003).
94. Wong, W.M. et al. Histogenesis of human colorectal adenomas and hyperplastic polyps: the role of cell proliferation and crypt fission. *Gut* **50**, 212-7 (2002).
95. van den Brink, G.R. & Offerhaus, G.J. The morphogenetic code and colon cancer development. *Cancer Cell* **11**, 109-17 (2007).

96. Cheng, H., Bjerknes, M., Amar, J. & Gardiner, G. Crypt production in normal and diseased human colonic epithelium. *Anat Rec* **216**, 44-8 (1986).
97. Jen, J. et al. Molecular determinants of dysplasia in colorectal lesions. *Cancer Res* **54**, 5523-6 (1994).
98. Takayama, T. et al. Analysis of K-ras, APC, and beta-catenin in aberrant crypt foci in sporadic adenoma, cancer, and familial adenomatous polyposis. *Gastroenterology* **121**, 599-611 (2001).
99. Jass, J.R., Whitehall, V.L., Young, J. & Leggett, B.A. Emerging concepts in colorectal neoplasia. *Gastroenterology* **123**, 862-76 (2002).
100. Otori, K. et al. Infrequent somatic mutation of the adenomatous polyposis coli gene in aberrant crypt foci of human colon tissue. *Cancer* **83**, 896-900 (1998).
101. Redston, M. Carcinogenesis in the GI tract: from morphology to genetics and back again. *Mod Pathol* **14**, 236-45 (2001).
102. Fialkow, P.J. Clonal origin of human tumors. *Biochim Biophys Acta* **458**, 283-321 (1976).
103. Vogelstein, B., Fearon, E.R., Hamilton, S.R. & Feinberg, A.P. Use of restriction fragment length polymorphisms to determine the clonal origin of human tumors. *Science* **227**, 642-5 (1985).
104. Fearon, E.R., Hamilton, S.R. & Vogelstein, B. Clonal analysis of human colorectal tumors. *Science* **238**, 193-7 (1987).
105. Merritt, A.J., Gould, K.A. & Dove, W.F. Polyclonal structure of intestinal adenomas in ApcMin/+ mice with concomitant loss of Apc+ from all tumor lineages. *Proc Natl Acad Sci U S A* **94**, 13927-31 (1997).
106. Shibata, D., Schaeffer, J., Li, Z.H., Capella, G. & Perucho, M. Genetic heterogeneity of the c-K-ras locus in colorectal adenomas but not in adenocarcinomas. *J Natl Cancer Inst* **85**, 1058-63 (1993).
107. Thirlwell, C. et al. Clonality assessment and clonal ordering of individual neoplastic crypts shows polyclonality of colorectal adenomas. *Gastroenterology* **138**, 1441-54, 1454 e1-7 (2010).
108. Axelrod, R., Axelrod, D.E. & Pienta, K.J. Evolution of cooperation among tumor cells. *Proc Natl Acad Sci U S A* **103**, 13474-9 (2006).
109. Thliveris, A.T. et al. Polyclonality of familial murine adenomas: analyses of mouse chimeras with low tumor multiplicity suggest short-range interactions. *Proc Natl Acad Sci U S A* **102**, 6960-5 (2005).
110. Ishiguro, K., Yoshida, T., Yagishita, H., Numata, Y. & Okayasu, T. Epithelial and stromal genetic instability contributes to genesis of colorectal adenomas. *Gut* **55**, 695-702 (2006).
111. Bian, Y. et al. Somatic acquisition of TGFBR1*6A by epithelial and stromal cells during head and neck and colon cancer development. *Hum Mol Genet* **16**, 3128-35 (2007).
112. Tsao, J.L. et al. Tracing cell fates in human colorectal tumors from somatic microsatellite mutations: evidence of adenomas with stem cell architecture. *Am J Pathol* **153**, 1189-200 (1998).
113. Vogelstein, B. et al. Genetic alterations during colorectal-tumor development. *N Engl J Med* **319**, 525-32 (1988).
114. Fearon, E.R. & Vogelstein, B. A genetic model for colorectal tumorigenesis. *Cell* **61**, 759-67 (1990).

115. Reya, T., Morrison, S.J., Clarke, M.F. & Weissman, I.L. Stem cells, cancer, and cancer stem cells. *Nature* **414**, 105-11 (2001).
116. Barker, N. et al. Crypt stem cells as the cells-of-origin of intestinal cancer. *Nature* **457**, 608-11 (2009).
117. Bjerknes, M. & Cheng, H. Multipotential stem cells in adult mouse gastric epithelium. *Am J Physiol Gastrointest Liver Physiol* **283**, G767-77 (2002).
118. Lamprecht, S.A. & Lipkin, M. Migrating colonic crypt epithelial cells: primary targets for transformation. *Carcinogenesis* **23**, 1777-80 (2002).
119. Makino, T. et al. Primary signet-ring cell carcinoma of the colon and rectum: report of eight cases and review of 154 Japanese cases. *Hepatogastroenterology* **53**, 845-9 (2006).
120. Vogelsang, H. & Siewert, J.R. Endocrine tumours of the hindgut. *Best Pract Res Clin Gastroenterol* **19**, 739-51 (2005).
121. Rubio, C.A., Kanter, L., Bjork, J., Poppen, B. & Bry, L. Paneth cell-rich flat adenoma of the rectum: report of a case. *Jpn J Cancer Res* **87**, 109-12 (1996).
122. Moolgavkar, S.H. & Luebeck, E.G. Multistage carcinogenesis and the incidence of human cancer. *Genes Chromosomes Cancer* **38**, 302-6 (2003).
123. Leggett, B. & Whitehall, V. Role of the serrated pathway in colorectal cancer pathogenesis. *Gastroenterology* **138**, 2088-100 (2010).
124. Kimura, M. Evolutionary rate at the molecular level. *Nature* **217**, 624-6 (1968).
125. Fisher, J.C. Multiple-mutation theory of carcinogenesis. *Nature* **181**, 651-2 (1958).
126. Cairns, J. Mutation selection and the natural history of cancer. *Nature* **255**, 197-200 (1975).
127. Nowell, P.C. The clonal evolution of tumor cell populations. *Science* **194**, 23-8 (1976).
128. Smith, J.M. & Haigh, J. The hitch-hiking effect of a favourable gene. *Genet Res* **23**, 23-35 (1974).
129. Kinzler, K.W. & Vogelstein, B. Lessons from hereditary colorectal cancer. *Cell* **87**, 159-70 (1996).
130. Lynch, H.T., Smyrk, T. & Lynch, J.F. Overview of natural history, pathology, molecular genetics and management of HNPCC (Lynch Syndrome). *Int J Cancer* **69**, 38-43 (1996).
131. de la Chapelle, A. Genetic predisposition to colorectal cancer. *Nat Rev Cancer* **4**, 769-80 (2004).
132. Powell, S.M. et al. APC mutations occur early during colorectal tumorigenesis. *Nature* **359**, 235-7 (1992).
133. Miyaki, M. et al. Characteristics of somatic mutation of the adenomatous polyposis coli gene in colorectal tumors. *Cancer Res* **54**, 3011-20 (1994).
134. Kane, M.F. et al. Methylation of the hMLH1 promoter correlates with lack of expression of hMLH1 in sporadic colon tumors and mismatch repair-defective human tumor cell lines. *Cancer Res* **57**, 808-11 (1997).
135. Weisenberger, D.J. et al. CpG island methylator phenotype underlies sporadic microsatellite instability and is tightly associated with BRAF mutation in colorectal cancer. *Nat Genet* **38**, 787-93 (2006).
136. Forrester, K., Almoguera, C., Han, K., Grizzle, W.E. & Perucho, M. Detection of high incidence of K-ras oncogenes during human colon tumorigenesis. *Nature* **327**, 298-303 (1987).

137. Bos, J.L. et al. Prevalence of ras gene mutations in human colorectal cancers. *Nature* **327**, 293-7 (1987).
138. Tomlinson, I.P., Novelli, M.R. & Bodmer, W.F. The mutation rate and cancer. *Proc Natl Acad Sci U S A* **93**, 14800-3 (1996).
139. Loeb, L.A. A mutator phenotype in cancer. *Cancer Res* **61**, 3230-9 (2001).
140. Loeb, L.A. Mutator phenotype may be required for multistage carcinogenesis. *Cancer Res* **51**, 3075-9 (1991).
141. Knudson, A.G., Jr. Mutation and cancer: statistical study of retinoblastoma. *Proc Natl Acad Sci U S A* **68**, 820-3 (1971).
142. Luebeck, E.G. & Moolgavkar, S.H. Multistage carcinogenesis and the incidence of colorectal cancer. *Proc Natl Acad Sci U S A* **99**, 15095-100 (2002).
143. Lamlum, H. et al. The type of somatic mutation at APC in familial adenomatous polyposis is determined by the site of the germline mutation: a new facet to Knudson's 'two-hit' hypothesis. *Nat Med* **5**, 1071-5 (1999).
144. Baker, S.J. et al. Chromosome 17 deletions and p53 gene mutations in colorectal carcinomas. *Science* **244**, 217-21 (1989).
145. Fodde, R., Smits, R. & Clevers, H. APC, signal transduction and genetic instability in colorectal cancer. *Nat Rev Cancer* **1**, 55-67 (2001).
146. Samowitz, W.S. et al. Evaluation of a large, population-based sample supports a CpG island methylator phenotype in colon cancer. *Gastroenterology* **129**, 837-45 (2005).
147. Ang, P.W. et al. Comprehensive profiling of DNA methylation in colorectal cancer reveals subgroups with distinct clinicopathological and molecular features. *BMC Cancer* **10**, 227 (2010).
148. Snover, D.C., Jass, J.R., Fenoglio-Preiser, C. & Batts, K.P. Serrated polyps of the large intestine: a morphologic and molecular review of an evolving concept. *Am J Clin Pathol* **124**, 380-91 (2005).
149. O'Brien, M.J. et al. Comparison of microsatellite instability, CpG island methylation phenotype, BRAF and KRAS status in serrated polyps and traditional adenomas indicates separate pathways to distinct colorectal carcinoma end points. *Am J Surg Pathol* **30**, 1491-501 (2006).
150. Kambara, T. et al. BRAF mutation is associated with DNA methylation in serrated polyps and cancers of the colorectum. *Gut* **53**, 1137-44 (2004).
151. Barker, N. & Clevers, H. Tumor environment: a potent driving force in colorectal cancer? *Trends Mol Med* **7**, 535-7 (2001).
152. Muto, T., Bussey, H.J. & Morson, B.C. The evolution of cancer of the colon and rectum. *Cancer* **36**, 2251-70 (1975).
153. Stryker, S.J. et al. Natural history of untreated colonic polyps. *Gastroenterology* **93**, 1009-13 (1987).
154. Jones, S. et al. Comparative lesion sequencing provides insights into tumor evolution. *Proc Natl Acad Sci U S A* **105**, 4283-8 (2008).
155. Welin, S., Youker, J. & Spratt, J.S., Jr. The rates and patterns of growth of 375 tumours of the large intestine and rectum observed serially by double contrast enema study (malmore technique). *Am J Roentgenol Radium Ther Nucl Med* **90**, 673-87 (1963).
156. Bolin, S., Nilsson, E. & Sjodahl, R. Carcinoma of the colon and rectum--growth rate. *Ann Surg* **198**, 151-8 (1983).

157. Tada, M., Misaki, F. & Kawai, K. Growth rates of colorectal carcinoma and adenoma by roentgenologic follow-up observations. *Gastroenterol Jpn* **19**, 550-5 (1984).
158. Hoff, G., Foerster, A., Vatn, M.H., Sauar, J. & Larsen, S. Epidemiology of polyps in the rectum and colon. Recovery and evaluation of unresected polyps 2 years after detection. *Scand J Gastroenterol* **21**, 853-62 (1986).
159. Hofstad, B. et al. Growth of colorectal polyps: redetection and evaluation of unresected polyps for a period of three years. *Gut* **39**, 449-56 (1996).
160. Bersentes, K., Fennerty, M.B., Sampliner, R.E. & Garewal, H.S. Lack of spontaneous regression of tubular adenomas in two years of follow-up. *Am J Gastroenterol* **92**, 1117-20 (1997).
161. Knoernschild, H.E. Growth rate and malignant potential of colonic polyps: early results. *Surg Forum* **14**, 137-8 (1963).
162. Johnson, D.A. et al. A prospective study of the prevalence of colonic neoplasms in asymptomatic patients with an age-related risk. *Am J Gastroenterol* **85**, 969-74 (1990).
163. DiSario, J.A., Foutch, P.G., Mai, H.D., Pardy, K. & Manne, R.K. Prevalence and malignant potential of colorectal polyps in asymptomatic, average-risk men. *Am J Gastroenterol* **86**, 941-5 (1991).
164. Hirano, M. et al. Apparent mtDNA heteroplasmy in Alzheimer's disease patients and in normals due to PCR amplification of nucleus-embedded mtDNA pseudogenes. *Proc Natl Acad Sci U S A* **94**, 14894-9 (1997).
165. Clement-Ziza, M., Munnich, A., Lyonnet, S., Jaubert, F. & Besmond, C. Stabilization of RNA during laser capture microdissection by performing experiments under argon atmosphere or using ethanol as a solvent in staining solutions. *RNA* **14**, 2698-704 (2008).
166. Keays, K.M., Owens, G.P., Ritchie, A.M., Gilden, D.H. & Burgoon, M.P. Laser capture microdissection and single-cell RT-PCR without RNA purification. *J Immunol Methods* **302**, 90-8 (2005).
167. Fink, L. & Bohle, R.M. Laser microdissection and RNA analysis. *Methods Mol Biol* **293**, 167-85 (2005).
168. Kim, K.M. & Shibata, D. Tracing ancestry with methylation patterns: most crypts appear distantly related in normal adult human colon. *BMC Gastroenterol* **4**, 8 (2004).
169. Segditsas, S. & Tomlinson, I. Colorectal cancer and genetic alterations in the Wnt pathway. *Oncogene* **25**, 7531-7 (2006).
170. Andreyev, H.J., Norman, A.R., Cunningham, D., Oates, J.R. & Clarke, P.A. Kirsten ras mutations in patients with colorectal cancer: the multicenter "RASCAL" study. *J Natl Cancer Inst* **90**, 675-84 (1998).
171. Stone, J.G., Tomlinson, I.P. & Houlston, R.S. Optimising methods for determining RER status in colorectal cancers. *Cancer Lett* **149**, 15-20 (2000).
172. Yang, A.S. et al. A simple method for estimating global DNA methylation using bisulfite PCR of repetitive DNA elements. *Nucleic Acids Res* **32**, e38 (2004).
173. Clayton, D.A. Replication and transcription of vertebrate mitochondrial DNA. *Annu Rev Cell Biol* **7**, 453-78 (1991).
174. Graham, T.A. et al. Use of methylation patterns to determine expansion of stem cell clones in human colon tissue. *Gastroenterology* **140**, 1241-1250 e1-9 (2011).

175. Marshman, E., Booth, C. & Potten, C.S. The intestinal epithelial stem cell. *Bioessays* **24**, 91-8 (2002).
176. Barker, N., van de Wetering, M. & Clevers, H. The intestinal stem cell. *Genes Dev* **22**, 1856-64 (2008).
177. Lonnerholm, G., Selking, O. & Wistrand, P.J. Amount and distribution of carbonic anhydrases CA I and CA II in the gastrointestinal tract. *Gastroenterology* **88**, 1151-61 (1985).
178. Bekku, S. et al. Carbonic anhydrase I and II as a differentiation marker of human and rat colonic enterocytes. *Res Exp Med (Berl)* **198**, 175-85 (1998).
179. Baris, O.R. et al. The mitochondrial electron transport chain is dispensable for proliferation and differentiation of epidermal progenitor cells. *Stem Cells* **29**, 1459-68 (2011).
180. Ushijima, T. et al. Fidelity of the methylation pattern and its variation in the genome. *Genome Res* **13**, 868-74 (2003).
181. Laird, C.D. et al. Hairpin-bisulfite PCR: assessing epigenetic methylation patterns on complementary strands of individual DNA molecules. *Proc Natl Acad Sci U S A* **101**, 204-9 (2004).
182. Park, H.S. et al. Effects of epidermal growth factor and dimethylhydrazine on crypt size, cell proliferation, and crypt fission in the rat colon. Cell proliferation and crypt fission are controlled independently. *Am J Pathol* **151**, 843-52 (1997).
183. Sunami, E., de Maat, M., Vu, A., Turner, R.R. & Hoon, D.S. LINE-1 hypomethylation during primary colon cancer progression. *PLoS One* **6**, e18884 (2011).
184. Tomlinson, I.P., Roylance, R. & Houlston, R.S. Two hits revisited again. *J Med Genet* **38**, 81-5 (2001).
185. Li, J., Mizukami, Y., Zhang, X., Jo, W.S. & Chung, D.C. Oncogenic K-ras stimulates Wnt signaling in colon cancer through inhibition of GSK-3beta. *Gastroenterology* **128**, 1907-18 (2005).
186. Janssen, K.P. et al. APC and oncogenic KRAS are synergistic in enhancing Wnt signaling in intestinal tumor formation and progression. *Gastroenterology* **131**, 1096-109 (2006).
187. Luo, F. et al. Mutated K-ras(Asp12) promotes tumourigenesis in Apc(Min) mice more in the large than the small intestines, with synergistic effects between K-ras and Wnt pathways. *Int J Exp Pathol* **90**, 558-74 (2009).
188. Moreno, E. & Basler, K. dMyc transforms cells into super-competitors. *Cell* **117**, 117-29 (2004).
189. Qiu, W. et al. Chemoprevention by nonsteroidal anti-inflammatory drugs eliminates oncogenic intestinal stem cells via SMAC-dependent apoptosis. *Proc Natl Acad Sci U S A* **107**, 20027-32 (2010).
190. Baron, J.A. et al. A randomized trial of aspirin to prevent colorectal adenomas. *N Engl J Med* **348**, 891-9 (2003).
191. Benamouzig, R. et al. Daily soluble aspirin and prevention of colorectal adenoma recurrence: one-year results of the APACC trial. *Gastroenterology* **125**, 328-36 (2003).
192. Logan, R.F. et al. Aspirin and folic acid for the prevention of recurrent colorectal adenomas. *Gastroenterology* **134**, 29-38 (2008).
193. Cole, B.F. et al. Aspirin for the chemoprevention of colorectal adenomas: meta-analysis of the randomized trials. *J Natl Cancer Inst* **101**, 256-66 (2009).

194. Bertagnolli, M.M. et al. Celecoxib for the prevention of sporadic colorectal adenomas. *N Engl J Med* **355**, 873-84 (2006).
195. Arber, N. et al. Celecoxib for the prevention of colorectal adenomatous polyps. *N Engl J Med* **355**, 885-95 (2006).
196. Rothwell, P.M. et al. Long-term effect of aspirin on colorectal cancer incidence and mortality: 20-year follow-up of five randomised trials. *Lancet* **376**, 1741-50 (2010).
197. Terpstra, O.T., van Blankenstein, M., Dees, J. & Eilers, G.A. Abnormal pattern of cell proliferation in the entire colonic mucosa of patients with colon adenoma or cancer. *Gastroenterology* **92**, 704-8 (1987).
198. Potten, C.S., Kellett, M., Rew, D.A. & Roberts, S.A. Proliferation in human gastrointestinal epithelium using bromodeoxyuridine in vivo: data for different sites, proximity to a tumour, and polyposis coli. *Gut* **33**, 524-9 (1992).
199. Potten, C.S., Kellett, M., Roberts, S.A., Rew, D.A. & Wilson, G.D. Measurement of in vivo proliferation in human colorectal mucosa using bromodeoxyuridine. *Gut* **33**, 71-8 (1992).
200. Becciolini, A. et al. Colonic cell proliferation in normal mucosa of patients with colon cancer. *Acta Oncol* **37**, 65-71 (1998).